Univerzita Karlova v Praze
Filozofická fakulta

Ústav anglického jazyka a didaktiky

Diplomová práce

Ondřej Slówik

Rytmické rozdíly mezi vietnamskou angličtinou a britským standardem

Rhythmic Differences between Vietnamese English and the British Standard

Praha, 2013 vedoucí práce: doc. PhDr. Jan Volín, PhD.

Prohlašuji, že jsem magisterskou práci vypracoval samostatně a že jsem uvedl všechny použité prameny a literaturu.
Souhlasím se zapůjčením magisterské práce ke studijním účelům.

V Praze, 2013

I declare that the following MA thesis is my own work for which I used only the sources and literature mentioned.
I have no objections to the MA thesis being borrowed and used for study purposes.

Děkuji doc. Janu Volínovi za cenné rady a vstřícnost při konzultacích týkajících se mé diplomové práce. Dále děkuji mgr. Lence Weingartové za pomoc s technickým zpracováním získaných dat. Velký dík patří také všem vietnamským mluvčím, kteří se uvolili poskynout mi nahrávky svého hlasu.

Language is my mother, my father, my husband, my brother, my sister, my whore, my mistress, my check-out girl. Language is a complimentary moist lemon-scented cleansing square or handy freshen-up wipette. Language is the dew on a fresh apple; it's the soft rain of dust that falls into a shaft of morning light as you pluck from an old bookshelf a half-forgotten book of erotic memoirs. Language is the creak on a stair, it's a spluttering match held to a frosted pane, it's a half-remembered childhood birthday party, it's the warm, wet, trusting touch of a leaking nappy, the underside of a granite boulder, the first downy growth on the upper lip of a Mediterranean girl. It's cobwebs long since overrun by an old Wellington boot.

Stephen Fry

Abstract

This thesis deals with rhythmic differences between Vietnamese English and the British Standard. As all the recorded speakers were originally from Northern Vietnam or lived there for an extensive period of time, it should be noted that the subject of analysis was in fact North Vietnamese English.

The theoretical part describes the concept of rhythm in general and modern approaches to its analysing and measuring. Furthermore, the theoretical chapter describes the basics of phonetic characteristics of the Vietnamese language and the British standard of English.

The last chapter of the theoretical part constitutes a bridge between theory and analysis as it explains selected features of Vietnamese English concerning mainly the realization of vowels and consonants.

The next part is dedicated to methodology and it informs the reader about the criteria for selection of speakers and the means of gathering and processing material. At the end, a number of hypotheses regarding Vietnamese English are presented.

In the analysis, values for rhythm metrics for Vietnamese English (%V, ΔV, ΔC, varcoV, varcoC, rPVI-V, rPVI-C, nPVI-V, nPVI-C) are calculated, compared to the rhythm metrics for British English and further evaluated in relation to gender, speakers and prosodic compactness.

The results of the analysis show that vocalic measurements of Vietnamese English are slightly lower than in the case of British English. Consonantal measurements, on the other hand, are markedly higher. Gender does not seem to be an influential factor in the rhythm of Vietnamese English. The analysis of prosodic compactness suggests that the more compact the speaker is, the more his/her speech resembles the British standard.

Keywords: *rhythm; British English, RP, Vietnamese English, phonetics*

Abstrakt

Diplomová práce se zabývá rytmickými rozdílnostmi mezi vietnamskou angličtinou a britským standardem. Vzhledem ke skutečnosti, že všichni nahrávaní mluvčí se narodili v severním Vietnamu, nebo tam aspoň dlouhodobě žili, je třeba specifikovat, že se jedná spíše o severovietnamskou angličtinu.

Teoretická část se zpočátku věnuje tématu rytmu z obecného hlediska a následně moderním metodám analýzy a měření rytmu. Teoretická část se dále věnuje základní charakteristice vietnamštiny a standardní britské angličtiny.

Poslední kapitola teoretické části představuje přemostění mezi teorií a analýzou vzhledem ke skutečnosti, že se zaobírá vybranými problémy vietnamské angličny, zejména konkrétními odlišnostmi při realizaci samohlásek a souhlásek.

Další kapitoly práce se zabývají metodologií výzkumu a vysvětlují kritéria výběru mluvčích a způsob zpracování materiálu.

Analýza se zabývá výpočtem parametrů rytmicity (%V, ΔV, ΔC, varcoV, varcoC, rPVI-V, rPVI-C, nPVI-V, nPVI-C) a jejich porovnání s britským standardem. V rámci analýzy byl též zkoumán vliv prozodické kompaktnosti a také pohlaví mluvčího na rytmus projevu.

Výsledky ukázaly, že vokalická měření u vietnamské angličtiny vyšla nepatrně nižší než u britského standardu. Hodnoty souhlásek však vyšly v případě vietnamské angličtiny do značné míry vyšší. Pohlaví mluvčího dle získaných výsledků nemá na řečový rytmus zásadní vliv. Na druhé straně hodnoty měření kompatních mluvčích se blíží hodnotám britské angličtiny, kdežto hodnoty méně kompaktních mluvčích se od britského standardu liší více.

Klíčová slova: *rytmus, vietnamská angličtina, britská angličtina, RP, fonetika*

Table of Contents

Abbreviations

BrE	British English
RP	Received Pronunciation
VE	Vietnamese English
BG	Breath group
ANOVA	Analysis of variance
V	Vowel
C	Consonant
%V	Percentage of vocalic duration within an utterance
ΔV	Standard deviation of vocalic intervals in an utterance
ΔC	Standard deviation of consonantal intervals in an utterance
rPVI	Raw Pairwise Variability Index
nPVI	Normalized Pairwise Variability Index
varco	Variation coefficient

1. Introduction

In spite of English becoming the global *lingua franca*, Latin of the nuclear age, it still retains its pedantic Higginsian accent consciousness. There are many more non-native speakers of English than there are native speakers. Communication in the English language nowadays involves native speakers only rarely in respect to the overall bulk of communication and yet students in classrooms all over the world often experience frustration trying to mimic the RP or GenAm accent in order to escape bullying from individuals whose ears and larynxes are better equipped for recognition and production of foreign speech sounds. On the other hand, some students are keen on attempting to imitate native accents while they seek praise from teachers and native speakers for having native-like annunciation. Furthermore, it is necessary to maintain some degree of phonetic standard for the learners to aspire to in order to ensure efficient communication. The only question remains whether the referential accent needs to be defined regionally as it is now or whether the students could and should learn some kind of "World English".

Vietnamese accent is perceived negatively in the English speaking world mainly because of low social status of the first generation Vietnamese immigrants determined by the nature of work they do (restaurants, shop keeping) and also because of their lack of interest in learning English and opening their community to the native cultural influence. In America, there are also other historical and political reasons for dislike towards the Vietnamese. Some perceive them as a painful memory of the war, others as immigrants and potential working opportunity thieves.

Among the international community in Vietnam the local accent is a frequent source of amusement. Many English native speakers live under the impression that the Asians do not try hard enough or even that they consciously produce incomprehensible speech in order to annoy their listeners. Such approach can be only seen as unjust, ignorant and most likely caused by their never having to learn any foreign language properly.

Most Asian and European languages are so different in terms of articulation that it renders many individuals who attempt to learn some of them in adulthood and without sufficient didactic know-how frustrated, discouraged and hence unable to succeed. The Western civilization can only consider itself extremely lucky that no Asian language has yet become a global means of communication and that people still do not leave the West in great numbers in order to seek better life in Asia that would require local linguistic competence.

The state of EFL education in South-East Asia varies regionally to a great extent. Due to the language policy, the most proficient speakers come from the Philippines, Singapore and Indonesia. English has the status of the official language in all three countries. In Thailand, Cambodia, Myanmar, Laos and Vietnam, the English language has the status of a main foreign language taught from the elementary school. Thai, Cambodian and Burmese accents seemed perceptually easier to comprehend than Laotian and Vietnamese accents. Explanation for the difference could be seen in the presence/absence of tonality. Cambodian is a non-tonal language whereas Laotian and Vietnamese possess large tone inventories. Thai and Burmese are also tonal languages, which does not conform to the hypothesis, but Burma was a British colony for an extensive period of time and the Burmese language contains a large number of English loanwords; Thailand (Siam) has never been a colony but a long-term British influence is still apparent in the Thai culture and architecture.

After having taught English and Czech to Vietnamese students I learnt that pronunciation tends to be the key problem of their foreign language acquisition. I taught a PhD student from the Faculty of Engineering whose vocabulary range was extensive and the command of English syntax was quite advanced but when he attempted to utter a sentence, I could understand him only with great difficulty and mainly because I was already "tuned in" the idiolect of Vietnamese learners of English. At that point I realized that the field of Vietnamese English, if harvested properly, could offer me an interesting topic for my MA thesis. Furthermore, provided that the analysis yields satisfactory results, further research could come up with a didactic tool that might help Vietnamese learners ease their struggle with the English pronunciation.

An accent constitutes a collage of segmental and supra-segmental features. This thesis attempts to analyze the differences in rhythm between Vietnamese English and the British RP standard. The choice of a supra-segmental topic was conditioned by two factors. First, some studies dealing with Vietnamese accent in English segmental level have already been published and they will be mentioned in section 2.5. Some of their findings will be illustrated by examples taken from our recordings. Second, there have been studies on a similar topic (differences between the rhythm in various languages and English) successfully published and therefore this work can draw inspiration from their methodology. Moreover, by using the same experimental text for the recording as it has been used in the research of the Department of Phonetics and Phonology, we will be able to create a corpus of yet another accent that will be comparable to the ones that are already available.

Four of the BBC texts available at the department of Phonetics were selected for the purposes of the recording. Twelve Vietnamese advanced speakers of English (7 women, 5 men; age 23-32) were recorded in Hanoi. The recordings were electronically processed and segmented (see Chapter 3) and then we proceeded to the measurements and comparison of results.

Before we start to describe and analyse the data, we need to address the theoretical aspects the comprehension of which is necessary for successful analysis of the gathered data. In the next chapter we will look at the concept of rhythm in general, then at the methods of measuring rhythm relevant to our research and finally we will discuss fundamental characteristics of the Vietnamese language, RP English and Vietnamese English.

2. Theoretical Background

The aim of this chapter is to offer theoretical foundations that will support the analysis and ensure that it is not based on false premises. At first, we are going to explain a few basic facts about rhythm and measuring of rhythm in speech. Then we are going to focus on Vietnamese, RP English and the Vietnamese accent in English.

2.1 Rhythm

When we think of rhythm, we closely associate it with sound and language but the truth is that rhythm is omnipresent in the world around us and it can be perceived by different senses than just hearing.

According to the OED definition, rhythm constitutes the "movement of procedure with uniform or patterned recurrence"[1] related to the Greek word **ῥυθμός** (rhythmós) meaning "measured motion, time whether in sound or motion" as used by Democritus or "measure, proportion or symmetry" as used by Plato[2]. Such definitions favour the statement mentioned above regarding the fact that rhythm does not necessarily have to be related merely to sound and language. Heartbeat is definitely an example of rhythm and when we place our hand on our chest we perceive heart rhythm by means of touch. Our planet's revolving around the sun and the subsequent change of day and night is also a manifestation of rhythm, this time perceived by sight.

From the popularity of dance, music and poetry, or physical exercises, which are all heavily rhythm-based activities it is apparent that people find pleasure, entertainment and art value in rhythmicity. Moreover, as it can be easily tested on infants, rhythm irregularity or distortion is perceived as something undesirable and even unpleasant.

[1] Oxford English Dictionary (www.oed.com)

[2] Liddel&Scott, *A Greek-English Lexicon*. New York: Oxford University Press, 1996.

2.1.1 Rhythm in artistic forms of language

Rhythm in language does not merely serve the communicative function. In certain special cases, it is used to convey artistic properties of language. Moreover, rhythm was one of the key means of preservation of artistic language in the era before the origin of writing. As rhythmical speech is easier to remember, all fairytales, legends, riddles etc. were transmitted from generation to generation in rhythmically regular form. Even Homer's Iliad and Odyssey were initially meant for oral preservation.

Nowadays, poetry is surely the branch of artistic language that is most distinctly rhythmicized. The rhythmical element in poetry is as important as the semantic one and sometimes even more important. A large part of its artistic value comes from the fact that there is a certain recurring rhythmical pattern.

The semantic element in prose usually tends to override rhythmicity but it does not mean that prose does not possess any rhythmical pattern. Jankovič (2002: 165) claims that rhythm "can merge in with semantic or syntactic segmentation of general speech to such an extent that leaves us unaware of its presence"[3].

2.1.2 Rhythm in spoken communication

It could be argued that the artistic forms of language are also means of communication but we thought it important to draw a distinction between these two based on the assumption that the primary purpose of communication is to transfer a piece of information from one individual to another whereas the primary purpose of the artistic forms is to create aesthetic value. By "communication" for the purposes of this study we mean a simple unmarked transfer of information.

Neurophysiologic research of the human brain has shown that neurons send out impulses of energy in rhythmical patterns rather than sending them out constantly. Even during perception of speech signal, the brain focuses on the speaker's rhythm. It tunes into the rhythm and then predicts the upcoming communicative element according to the rhythmical pattern. As long as the speaker conforms to the

[3] Translation O.S.

established rhythm pattern, the listener is more effective and does not have to waste additional energy in order to understand the speaker. However, if the speaker deviates from his rhythm pattern or if the speaker's rhythm pattern is broken and inconsistent, the listener tends to become annoyed and disinterested. (Ghitza&Greenberg 2009) Such annoyance renders the communication less effective and less information is transferred and subsequently remembered. Rhythm deviation is a frequent cause of linguistic inefficiency of non-native speakers.

2.1.2.1 Components of language rhythm

According to Dauer:

> *The particular rhythm of a1anguage is the result of the interaction of a number of components, including phonetic components, such as the relative length, pitch., and segmental quality of accented and unaccented syllables, and phonological components, such as syllable structure and the function of accent. A system of rating whereby these components are broken down into features which can be assigned a plus or minus va1ue allows us to compare the rhythm of languages or language varieties. Languages which have "strong stress" or which have been labelled "stress-timed" are seen to share certain features. Rhythm is a tota1 effect involving phonetic and phonological as well as segmental and prosodic phenomena.* (Dauer 1987: 448)

She distinguishes 4 categories of the components: a) **length** – that is subdivided into *duration* (i.e. accented[4] syllables can be longer than unaccented syllables in some languages whereas in others the accent does not make any difference), *syllable structure* (i.e. some languages present a variety of heavy and light syllable types where heavy syllables tend to be accented more often; in other language types, the weight and accent of a syllable are independent) and *quantity* (i.e. some languages

[4] Dauer points out that she uses the term "accent" as the phonological feature which when realized promotes the perception of one particular syllable (or mora) in relation to others. Accent can then serve as a basis of rhythmic grouping. She reserves the term "stress" for the phonetic realization of certain kinds of linguistic accents.

only allow for vowel quantity distinction in accented syllables, other languages do not condition the restrictions of quantity by accents; b) **pitch** – subdivided into *intonation* (accented syllables can be turning points in the intonation contour and pitch can correlate with accent; in other cases, pitch and accent are independent) and *tone* (some tonal languages only exhibit tones on accented syllables, other syllables are atonal; other tonal languages operate with various sandhi rules that neutralize or alter tones on unaccented syllables and tones are fully developed merely on accented syllables; yet another group of tonal languages among which we can count also the Vietnamese language develops tones on unaccented as well as on accented syllables); c) **quality of vowels and consonants** – some languages retain the full vowel inventory in accented syllables only and vowels in unaccented syllables tend to be reduced or centralized, other languages do not employ reduction or centralization of vowels and if they do, it affects both accented and unaccented syllables. Consonants can be also articulated more carefully in accented syllables in certain languages and some can have special allophones (e.g. syllabic consonants). In other languages, however, articulation of consonants remains the same regardless of the syllable.
d) **function of accent** – Accent can be either fixed in one position or it can be movable in which case it has the power to change the meaning or create a new word.

2.1.2.2 Isochrony

The word isochrony is a combining form of Greek **ισος** (isos) – "equal in size, strength, or number" and **χρονος** (chronos) – "time"[5]. In 1940, in his book *Speech Signals in Telephony*, Arthur Loyd James divided languages into two categories according to rhythm for the first time. First category was labelled "Machine-gun rhythm" and the second "Morse code rhythm". The first category corresponds to the modern label "syllable-timed", i.e. isochrony occurs in between syllables, vowel reduction does not take place and all syllables display perceptionally roughly the same duration (French, Czech, Vietnamese) whereas the second category corresponds to the label "stress-timed", i.e. ischrony occurs between stresses, vowel

[5] Liddel&Scott, *A Greek-English Lexicon*. New York: Oxford University Press, 1996.

reduction takes place and the inter-stress segments tend to have roughly the same duration (English, Russian, Arabic). Abercombie explains the relation of languages and isochrony as follows:

> *As far as is known, every language in the world is spoken with one kind of rhythm or with the other. In the one kind, known as a syllable-timed rhythm, the periodic recurrence of movement is supplied by the syllable-producing process (...) the syllables recur at equal intervals of time - they are isochronous. (...) In the other kind, know as a stress-timed rhythm, the periodic recurrence of movement is supplied by the stress-producing process (...) the stressed syllables are isochronous.* (Abercombie, 1967:97)

It should be noted that the duration of the segments is measured in "mental time", i.e. subjectively perceived time that is impossible to measure by objective measuring methods.

2.1.2.3 P-centre

The phenomenon of p-centre (perceptual centre) arises from the attempts to seek for the perceptually relevant moments when our brain notices the appearance of the individual syllables. Our brain does not seem to be able to notice the syllable immediately; hence the p-centre is not likely to be placed at the very beginning of the syllable. Its actual location has not yet been agreed on but it is assumed for the initial third of the syllable.

2.2 Measuring speech rhythm

In the previous chapter we described the theoretical ground of rhythm including brief notes from the history of studying and researching rhythm. This chapter focuses on introducing the parameters that will be used for measuring rhythmical differences between Vietnamese accented English and the RP standard. The parameters in question are: %V; ΔV; ΔC; varcoV; varcoC; rPVI-V; rPVI-C; nPVI-V and nPVI-C.

It must be noted, however, that all these parameters are only correlates of rhythm and not rhythmical features as such hence the results yielded by using them must be handled with caution.

2.2.1 Interval measures (%V; ΔV; ΔC)

Ramus et al. (1999) introduced three rhythm metrics: a) *%V* – the proportion of total utterance duration comprising vocalic interval. b) *ΔV* – the standard deviation of vocalic interval. c) *ΔC* – the standard deviation of consonantal interval. Ramus et al. analysed eight languages and attempted to establish whether they were stress-timed, syllable-timed or mora-timed. They used the three metrics as decisive criteria and claimed that out of the eight languages English, Dutch and Polish were assessed to be stress-timed, French, Spanish, Italian and Catalan syllable-timed and Japanese mora-timed. The three classes showed different scores in %V and ΔC. What is important for the purposes of our study is that Ramus et al. discovered the tendency that stress-timed languages tend to have lower values of %V whereas syllable-timed languages tend to have lower values of ΔC.

2.2.2 Pairwise Variability Indices (rPVI-V; rPVI-C; nPVI-V; nPVI-C)

Grabe and Low (2002) devised the Pairwise Variability Index as a tool to measuring duration difference between successive syllables (specifically stressed and unstressed vowels) the presupposition being that the difference should be greater in stress-timed languages. There are two types of PVIs: a) raw (rPVI); b) normalized (nPVI)

a) According to most studies, Raw Pairwise Variability Index is suitable for measuring consonantal intervals because the differences in consonantal interval duration are assumed to be a phonotactic function. Therefore, they should be captured rather than normalised. The rPVI is calculated as the mean of the differences between successive intervals.

$$\text{rPVI} = \left(\sum_{k=1}^{m-1} |d_k - d_{k+1}| \right) \Big/ (m-1).$$

Fig. 2.1. *Formula for rPVI (m represents the number of intervals, d stands for the duration of the k^{th} interval).*

b) Normalised Pairwise Variability Index is supposed to be suitable for measuring vocalic intervals as it normalises the speech rate. nPVI is calculated as the mean of differences between successive intervals divided by the sum of the same intervals.

$$\text{nPVI} = 100 \times \left(\sum_{k=1}^{m-1} |(d_k - d_{k+1})/((d_k + d_{k+1})/2)| \right) \Big/ (m-1),$$

Fig. 2.2. *Formula for nPVI (m stands for the number of intervals, d represents the duration of the k^{th} interval).*

Despite the fact that most of the studies dealing with rhythm only compare rPVI-C and nPVI-V we will not disregard the results yielded by rPVI-V and nPVI-C because it is possible that they can reveal additional information in regard to the rhythm of Vietnamese English.

2.2.3 Variation coefficient ΔV and ΔC

Researchers after Ramus began to investigate the influence of speech rate on the measurements of ΔV and ΔC and established that speech rate could corrupt the measurements. Inspired by nPVI devised by Grabe and Low, Dellwo and Wagner (2003) came up with a means of normalising ΔV and ΔC and called it *varcoC*; *varcoV*. The figures are calculated as the standard deviation of consonantal/vocalic interval duration divided by the mean consonantal/vocalic duration and multiplied by 100. Dellwo and Wagner also found only a little correlation between %V and speech rate and thus they claimed that normalisation of %V was unnecessary.

2.3 Vietnamese

Vietnamese is a language spoken by approximately 92 million[6] people in Vietnam. According to a census[7] from the year 2000, there were over 1 million speakers of Vietnamese living in the US. Other substantial Vietnamese communities are in Australia (mainly people who escaped South Vietnam after 1975), France (descendants of people who came to France during the colonial era or shortly after) and the Eastern Bloc - Russia, the Ukraine, Czechoslovakia, Poland (inhabitants of North Vietnam coming during and after the Vietnamese-American war).

2.3.1 Official Language of Vietnam

Vietnam comprises 54 – 60 ethnicities (Slavická, 2008: 13). The Vietnamese language which presently occupies the position of the official language of Vietnam was originally spoken only by the ethnicity Việt settled in the river deltas and along the sea shore. Nowadays, the Việt people take up 88% of the overall Vietnamese population and the remaining 12% is distributed among the remaining 53+ ethnicities most of which still retain their own culture and language. Being a member of the leading ethnicity is still considered prestigious but, as opposed to the past, the ethnic minorities are not oppressed or forced to assimilate. They can freely cultivate their traditions, language and architecture. The government's approach to the minorities changed in the late 1990's when, mainly due the increase of tourism, it became apparent that retaining cultural differences would bring more profit than destroying them.

In terms of orthography, Vietnamese is one of South-East Asian languages using a writing system based on Latin script. The script is called *Chữ Quốc Ngữ* (National script) and it employs many Portuguese diacritical signs to capture tonal differences

[6] CIA factbook - https://www.cia.gov/library/publications/the-world-factbook/geos/vm.html

[7] http://www.census.gov/population/cen2000/phc-t20/tab05.pdf

and vowel quantity and quality. *Chữ Quốc Ngữ* is the third type of script used in Vietnam. Chinese characters became the first official writing system in Vietnam but they were subsequently replaced by *Chữ Nôm* (Southern Characters). *Chữ Quốc Ngữ* was invented in the Middle Ages by European missionaries but it was only after WWII when the script gained its official status.

2.3.2 Genealogy and typology

The French sinologist Henri Maspéro included Vietnamese into the Thai language family based on the claim that the tonal systems of both languages are similar and there are many Thai words present in the Vietnamese word-stock. However, in the mid 20th Century, another French linguist, André Haudricourt, introduced a more plausible classification that remains valid until the present times. The Vietnamese language is agreed to be a member of the Viet-Muong group, Mon-Khmer branch of the Austro-Asiatic language family.

It is estimated (Trần Trí Dõi, 2011) that about 60% of the Vietnamese vocabulary can be traced to Chinese but the languages are not genealogically closely related. Very often there are two words for one concept, one of which is considered purely Vietnamese and the other Sino-Vietnamese. Sino-Vietnamese words tend to be used in higher registers (art, poetry, rituals) whereas the purely Vietnamese words are reserved for everyday conversation. The Vietnamese words for *wind* and *water*, for instance, are ***gió*** and ***nước***, the Sino-Vietnamese ***phong*** and ***thuỷ***. *Phong thuỷ* is the Sino-Vietnamese transliteration of the Chinese characters 堪輿, which the Western world knows it as Feng-shui. Vietnamese speakers would never use the words *phong* or *thuỷ* when talking about weather. Analogically, they would not call Feng-shui *gió nước*.

Typological classification of Vietnamese in the Czech sources oscillates between the polysynthetic and isolating type (Skalička 2004, Čermák 2004). However, the extent of accuracy of such classification remains questionable as the matter of typological classification of Vietnamese is still understudied.

In Vietnamese there is no declension or conjugation; differences between word classes are not marked. Comparison of adjectives and temporal relations are expressed by adding a special particle to the adjective and verb respectively. A large portion of the Vietnamese vocabulary exhibits the pattern syllable = word and there are no lexemes longer than 3 syllables (with the exception of some modern terms from chemistry or other sciences). Some words can be perceived as longer but the underlying structure is the one of a head and a modifier e.g. *người bán hàng* (person – sell – goods) = seller. Traces of morphology can be found in word-formation e.g. the structure *chủ nghĩa* carries the meaning of *–ism* (*chủ nghĩa xã hội* – socialism; *xã hội* means *society* or *social*) and cannot stand alone.

2.3.3 Dialects

Vietnamese linguists (e.g. Hoàng Thị Châu, 1989) generally distinguish three basic dialects – Northern, Central and Southern. Thompson (1965) uses the same distinction only labelling them Tonkinese, Annamese and Cochinchinese (French colonial names of the corresponding areas). However, Thompson further distinguished the dialects of Hanoi, Saigon, Vinh, Hue and Danang. By doing so he pointed out that urban language in Vietnam is rather specific and can differ from the surrounding rural areas to a great extent.

Some dialects (especially in central Vietnam) differ from the Hanoian standard to such an extent that it brings them on the verge of intelligibility and the national TV must use subtitles for them. Vietnamese dialects differ lexically to a great extent. In terms of pronunciation they differ mainly in tone inventory and pronunciation of consonants. The orthography, however, remains unaltered.

2.3.3.1 Standard dialect

Since Vietnamese is so rich in terms of dialect, the question arose which of the dialects should be taken as referential. Until 1945, it was the dialect of the imperial city of Huế where the emperors of the last dynasty (called Nguyễn) were seated. At the time, Huế was the cultural centre of Vietnam. After the August revolution of 1945, Hanoi was chosen as the capital of the country. In 1954, when the French were driven away and the country divided, Hanoi was established as the capital of Northern Vietnam and in 1976 when the country reunited, Hanoi became the capital of the whole country again and Saigon, the capital of the South, was renamed Ho Chi Minh's City in the attempt to discourage those who still believed in the autonomy of the South. The dialect of Hanoi was also adopted as the standard dialect and it is rather apparent that it was more of a political decision than historical or linguistic one.

The choice of the Hanoian dialect might be linguistically justified on the grounds of tonality. The dialect of Hanoi (and most of other Northern dialects) exhibits the full inventory of six tones. In Saigon it is only five and in certain areas of central Vietnam merely four. On the other hand, Hanoian dialect is more homophonous than the others. The graphemes x|s; r|d|gi and ch|tr and ưu|ươu|iêu are not distinguished in speech and pronounced as [s]; [z]; [tʃ] and [i͡ɛ̯] whereas in the South the variation of pronunciation is retained.

Since all the speakers recorded in the present study were born in the North or they have been living there for extensive periods of time that ameliorated their vernacular, from now on, when we speak about the Vietnamese language we will consider the Northern dialect unless stated otherwise.

2.3.4 Phones

2.3.4.1 Monophtongs

Hữu Quỳnh - Vương Lộc (1980: 38) claim the standard number of Vietnamese vowels to be 11. They distinguish front vowels (/i/; /e/; /ɛ/), back unrounded vowels (/ă/; /ɤ̆/; /ɑ/; /ɤ/; /ɯ/) and back rounded vowels (/u/; /o/; /ɔ/). From the presence of **/ă/; /ɤ̆/** x **/ɑ/; /ɤ/** and **/o/; /ɔ/** x **/e/; /ɛ/** it can be deduced that Vietnamese distinguishes vowels based on quantity as well as quality. Only vowels can form the centre of a syllabic peak in Vietnamese. In some syllables the peak is preceded by the pre-tonal semi-vowel /u̯/ that is classified as the approximant /w/ in more recent works (Kirby 2011).

2.3.4.2 Diphthongs

Vietnamese in general accommodates three centring diphthongs (/i͡e̯/; /u͡o̯/; /ɯ͡ɤ̯/). As it was mentioned in 2.3.5.1, the standard dialect tends to pronounce the diphthong /ɯ͡ɤ̯/ as /i͡e̯/. The speakers themselves call it "simplification" of pronunciation because they perceive /ɯ͡ɤ̯/ rather difficult from the articulatory perspective. Kirby (2011) uses slightly different categorization of diphthongs from the traditional approach (Hữu Quỳnh - Vương Lộc 1980, Thompson 1965) employing the symbol for "schwa". The English "schwa" is roughly equivalent to the Vietnamese /ɤ̆/.

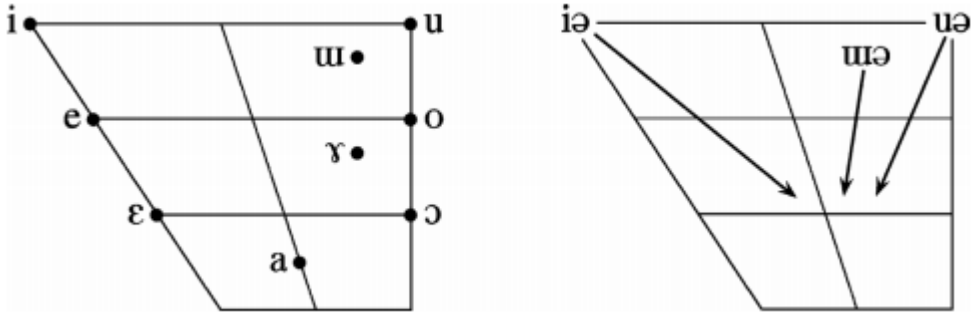


Fig. 2.3. *Positions of Northern Vietnamese vowels and diphthongs (Kirby 2011), short variants /ă/ and /ɤ̆/ of vowels /a/ and /ɤ/ are not listed probably because they*

are identical in terms of quality. Diphthongs finish in schwa as opposed to /e/; /o/ and /ɤ/, which is the traditional description in Vietnamese sources.

2.3.4.3 Consonants

There are no syllabic consonants or consonant clusters in Vietnamese and thus a consonant can only occur syllable-initially or syllable-finally. Furthermore, it is always followed or preceded by a vowel.

According to Thompson (1965) there are 18 initial consonants in the standard dialect: /tʰ/; /t/; /tʃ/; /k/; /ɓ/; /ɗ/; /m/; /n/; /ŋ/; /ɲ/; /f/; /s/; /x/; /h/; /v/; /f/; /ɣ/ (sometimes realized allophonically as /g/); /l/ (initial /p/ appears in a limited number of lexical borrowings but only rarely and so /p/ is not considered a "native" sound). Kirby (2011) reclassifies the semi-vowel /u̯/ to a labial approximant /w/ and lists it under consonants. Furthermore, he mentions the glottal stop that appears very frequently but it carries no semantic relevance.

	Labial	Labio-dental	Dental	Alveolar	Palatal	Velar	Glottal
Plosive	ɓ		t tʰ	ɗ	tɕ	k	ʔ
Nasal	m		n		ɲ	ŋ	
Fricative		f v		s z		x ɣ	h
Approximant	w						
Lateral approximant			l				

Fig. 2.4. *List of initial consonants in Northern Vietnamese (Kirby 2011).*

The choice of final consonants is limited to 8: four nasals /m/; /n/; /ŋ/; /ɲ/ and four voiceless plosives /p/; /t/; /k/; /ʈ/. Articulation of voiceless plosives does not involve any audible explosion, which renders their perceptual distinction problematic. Furthermore, syllables with final plosives only occur with tones 5 and 6 (see 2.3.6.1) and the tonal vowel appears to be perceptionally shorter due to the abrupt change of pitch before the final consonant. Kirby (2011) distinguishes final approximants /w, j/ that are classified as semi-vowels /u̯, i̯/ by Vietnamese phoneticians.

	Labial	Alveolar	Palatal	Velar
Nasal	m	n	ɲ	ŋ
Plosive	p	t	ʈ	k

Fig. 2.5. *List of final consonants in Northern Vietnamese.*

Many speakers of the Hanoian dialect (and some other Northern dialects) have difficulties distinguishing /l/ and /n/ in the syllable-initial position. The speakers occasionally utter /l/ when they mean /n/ or vice versa. For example, the word *lo* means "to be afraid" and the word *no* means "full, unable to eat any more". Deciphering the correct meaning must be done contextually and in regard to the particular speaker, i.e. we realize rather quickly that the speaker suffers from this condition and we start guessing the alternative meanings. In Southern dialects this problem occurs very scarcely. This phenomenon is not directly connected to rhythm but it is relevant nevertheless as such speakers are unable to distinguish these two consonants in English as well (see 2.5.2.5)

2.3.5 Tones

At the beginning of AD, no tones were present in the Vietnamese language whatsoever. The variety of speech sound combination was broader due to the possibility of consonant clusters. By the 6th century AD, all final consonant clusters were lost and replaced by three tones (level, falling, rising). The tonal inventory of modern Vietnamese language was reached in the 12th century after the loss of initial consonant clusters (Trần Trí Dõi, 2011: 87).

Traditional description of modern standard Vietnamese tonal inventory mentions six tones. Their Vietnamese labels are very helpful because they carry the tones they refer to (e.g. *huyền* is the name of the second tone while the syllable itself carries it) but an alternative numeral labelling had to be devised for people who lack the knowledge of the Vietnamese language. In older studies (Thompson 1965, Hữu

Quỳnh - Vương Lộc 1980) as well as in pedagogical texts, Vietnamese tones are distinguished solely on the basis of pitch contour. In modern studies (Pham 2003, Brunelle 2003, 2009, 2012), however, it was discovered that the tones are also distinguished based on voice quality.

Unlike in Chinese, there is no tonal sandhi in Vietnamese, which means that tones do not change according to their tonal environment. In order to clarify the phenomenon of tonal sandhi, we can use the example of the Chinese greeting 你好 comprised of syllables [ni^{3}] and [hao^{3}] in fact pronounced [ni^{2}hao^{3}] because the rule of sandhi in Chinese dictates that in case of two consecutive thirds tones, the first changes into the second. (Yip, 2002)

Vietnamese tones are labelled *ngang* (1), *huyền* (2), *ngã* (3), *hỏi* (4), *sắc* (5), *nặng* (6). We can notice that each tone is represent by a special diacritics (or its absence); moreover, most of the syllables have a primary lexical meaning and their use as tone labels is only secondary. *Ngang* means horizontal, *ngã* – to fall down, *hỏi* – to ask, *sắc* – sharp and *nặng* – heavy. *Huyền* is nowadays used only as a label for the tone or a first name for women but it used to carry the meaning of "sad".

The older phonetic approach distinguishing tones merely by pitch contour uses upper-index numbers that follow the syllable and reflect its pitch contour: *ngang* **[ŋɑŋ^{44}]**; *huyền* **[hu̯i̯e̯n^{32}]**; *ngã* **[ŋɑ^{3ʔ5}]**; *hỏi* **[hɔi̯^{213}]**; *sắc* **[săk^{35}]** *nặng* **[năŋ^{3ʔ1}]**. It should be pointed out that there is no consensus on the numbering (and glottalization marking) and so the labels can change with different authors.

Pham (2003) challenged the traditional approach claiming that pitch contour is merely one of the distinctive features and, moreover, there are pairs of tones that are impossible to distinguish based plainly on pitch contour.

Drawing inspiration from Pham (2003) and Michaud (2004), Brunelle (2009) formulates a classification more suitable for phonetic discourse. He labels *ngang* as A1; *huyền* as A2; *sắc* as B1; *nặng* as B2; *hỏi* as C1 and *ngã* as C2.

2.3.5.1 Checked syllables

Open syllables and syllables with final nasals can carry any of the six tones. Checked syllables, i.e. syllables closed by voiceless stops (Brunelle, 2009), can only carry sắc (B1) or nặng (B2). In older studies and didactic texts these tones are considered allophones in complementary distribution. Pham (2003) claims that they are actually two different tones and hence standard Vietnamese has an inventory of 8 tones. Contemporary researchers adopted this view but didactic texts still strictly follow the pitch-based 6-tone classification. The checked syllables constitute a major source of articulatory interference (see. 2.5.1.2) when Vietnamese speakers use English.

2.3.6 Syllabic structure

The only obligatory element in a Vietnamese syllable is the vowel serving as the syllabic peak. Onset, pre-tonal semi-vowel and coda are optional elements. The number of syllables containing all the elements at once is limited. The figures below show a scheme of Vietnamese syllables suggested by Hữu Quỳnh - Vương Lộc (1980: 41) and two concrete realizations of such syllables.

	TONE		
	RHYME		
onset	pre-vowel	peak	coda

Fig. 2.6. *General scheme of a Vietnamese syllable.*

nặng (B2)			
	u͡iệ͡t		
ŋ	u	ɪ͡e̯	t

hỏi (C1)			
	ɤ		
0	0	ɤ	0

Fig. 2.7. *Realizations of syllables nguyệt (moon) and ở (to be, to live).*

2.3.7 Intonation

Intonation in non-tonal languages can be described as "the ensemble of pitch variations in the course of an utterance" (Nooteboom 1999: 2) or "the product of a conflation of different prosodic systems of pitch contrast" (Crystal 1964: 6). Such descriptions, however, are valid only for languages in which pitch variation does not affect the lexical meaning. Pitch variation in Vietnamese is a decisive factor in respect to word meaning. The syllable *dưa* accompanied by a falling tone (*dừa*) means *a coconut*, the same syllable accompanied by a rising tone (*dứa*) means *a pineapple*.

Brunelle et al. (2012) mention that some studies have been carried out with the intention to establish the nature of lexical and sentential intonation in Vietnamese.
Declarative sentences are found to have a slight overall f0 declination. Interrogatives are described as having a high overall range, or a high range and a rise starting much before the sentence final question marker. (...) Imperatives are also described as having a high overall f0, possibly with an additional final rise and longer duration. (Brunelle et al. 2012:6)
Sentential intonation seems to be more prominent in faster or emotionally affected speech. In slow or careful speech, however, lexical intonation overrides it.

A possible reason for sentential intonation's being less salient in Vietnamese than in non-tonal languages is its relative redundancy. The Vietnamese language employs optional final particles in marking communicative functions: particles *hả* or *à* mark yes/no questions (*Người đó là sinh viên của anh **à**?* Is that person your student?); *đi* or *nhé* mark the imperative (*Đóng cửa **đi**!* Close the door!); *nhỉ* or *nhé* mark expected agreement (*Quyển sách này hay **nhỉ**?* This book is interesting, isn't it?); *chứ* or *mà* mark contradiction (*Đây là chị tôi **mà**!* But this IS my older sister!) etc.

Apparently, many functions (including pragmatic ones) that are expressed through intonation in non-tonal languages are expressed lexically in Vietnamese, which renders sentential intonation in Vietnamese less significant. Nonetheless, according to Brunelle et al. (2012), there are measurable differences in Vietnamese sentential intonation. The core of the issue lies in the fact that despite being measurable, they

seem to be negligible from the perspective of human perception and hence insignificant for real-world communication.

2.3.8 Stress

As Vietnamese is largely a monosyllabic language and many syllables represent whole semantic lexemes, placement of lexical stress here is rather straightforward as there are no more syllables to choose from. The amount of disyllabic lexemes, however, is still rather high. Stress placement in disyllabic lexemes is not strictly defined but the tendency seems to favour placing the stress on the second syllable: *sinh **viên*** - student; *bưu **điện*** – post office (Slavická 2008) although in some cases it can be placed on the first syllable (***tổ** chức* – organize, organization; ***tham** gia* – take part in sth.). Lexemes with more than two syllables are extremely rare. There are structures such as *người bán hàng* – seller but *người* is a human classifier, *bán* is the verb "to sell" and *hàng* means "goods" and, despite the fact that translates as one lexeme into English, it probably should not be treated as one lexeme in Vietnamese. Classification of Vietnamese lexical units still has not been agreed on.

Cunningham (2009) claims Vietnamese to be a syllable-timed language but she does not support this claim by any secondary sources. The rhythm metrics for Vietnamese have not been yet calculated hence the results of VE cannot be assessed in respect to the rhythmical nature of Vietnamese.

Sentential stress is dealt with in Cao Xuân Hạo (2007). Cao claims that sentential stress is used for distinguishing individual syntagmata as it is often difficult to establish which elements in the sentence belong together due to lack of overt word-class markers or inflection. Cao uses the example of the sentence: ***Lan**//đi mua **cá**// mí lị **khế**// về nấu **canh***. (**Lan** went to buy **fish** as well as **star fruit** then she returned home to make **soup**.). The individual syntagmata are separated by (//) and the stressed syllables are written in bold.

Thompson (1965: 106-107) distinguishes three types of stress: a) *heavy* – singles out the syllable or syllables of each pause group which carry the heaviest burden of conveying information. b) *weak* – accompanies syllables which bear the lowest information conveying load. They often refer to things which have been brought up earlier or which are expectable in the general context. c) *medium* – all remaining syllables.

Sentential stress is usually not carried by certain grammatical units (pronouns, prepositions, classifiers) but it is by others (vocative elements, temporal elements, intensifiers). In short sentences and phrases, all constituents can be stressed e.g. ***người cao*** – tall people (or people are tall); ***chó chạy*** – dogs run. Some disyllabic units can have both syllables stressed such as ***vợ chồng*** – married couple. *Vợ* means "wife" and *chồng* means "husband" when the two words stand alone, which the reason why both constituents are stressed whereas in the case of *sinh **viên*** (student) the two syllables cannot stand separately and hence the single stress on the second syllable.

Vietnamese, as opposed to Chinese, does not possess atonal syllables. The loss of tone in some Chinese syllables originates from the fact that they are unstressed. Unstressed syllables in Vietnamese tend to be shorter and the tone contour is far from canonical but tone is still present.

2.4 Standard British English

Since this study deals with the nature of Vietnamese accent in English, we had to select a standard in order to possess a referential accent to the Vietnamese speakers. Vietnamese people usually learn the American or Australian variant of English because they are geographically close to Australia and there is an easy access to American television in Vietnam. Nonetheless, we decided to use the British standard mainly because of technical reasons as the Phonetic Department has done rather extensive research focusing on precisely this accent.

According to Peter Roach:

> *The accent has been known for nearly a century as RECEIVED PRONUNCIATION, or by its abbreviation, RP. Early in the 20th century, Daniel Jones, the great exponent of the description of English pronunciation, named it PUBLIC SCHOOL PRONUNCIATION, but later changed the name to Received Pronunciation. Other names have been proposed, such as GENERAL BRITISH (GB) and EDUCATED SOUTHERN BRITISH ENGLISH. The present author's own preference is for the name BBC PRONUNCIATION or BBC ACCENT.* (Roach 2004: 240)

2.4.1 Monophtongs

RP English distinguishes 12 vowels (although the /ə/ has no standard representation in writing), as opposed to Vietnamese that distinguishes only 11. There are short vowels (e.g. /ɪ/) and long vowels (e.g. /iː/) present in RP English and therefore, on the superficial level, it seems that the vowels are distinguished based on both quality and quantity. However:

> *It should be understood that the terms 'long' and 'short' should be seen in relative terms: the vowels of both classes are subject to the lengthening and shortening effects found in English, with the result that a 'short' vowel may, in some contexts, be longer than a 'long' vowel in a different context. The length mark ː is used to mark the long vowels, though this is actually redundant since the vowel symbols already successfully distinguish each vowel from every other.* (Roach 2004:241)

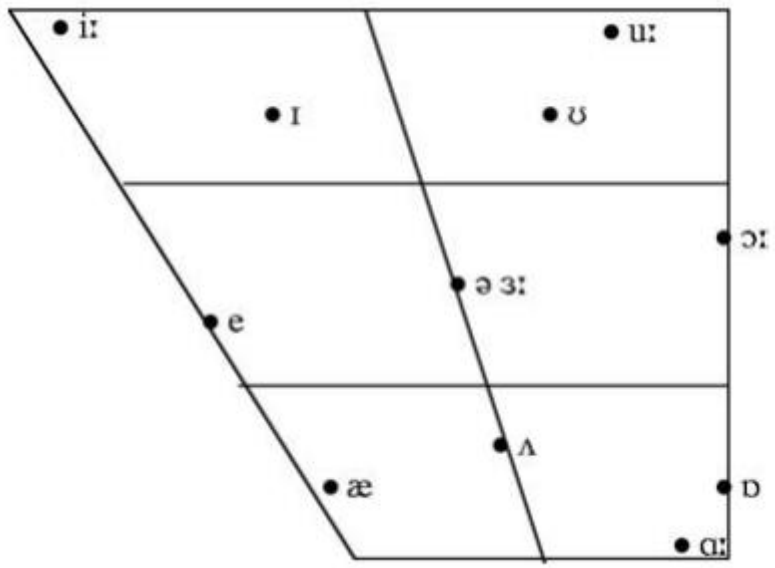


Fig. 2.8. *Monophtongs in RP English (Roach 2004).*

2.4.2 Diphthongs

In RP English, there are "centring" diphthongs ending with /ə/ and "closing" diphthongs ending with /ʊ/ or /ɪ/. Vietnamese only has centring diphthongs. English closing diphthongs can be accompanied by a final /ə/ and then they turn into phonetic triphthtongs (e.g. /f*aɪə*/). Triphthongs are quite rare and often smoothed or pronounced in two syllables.

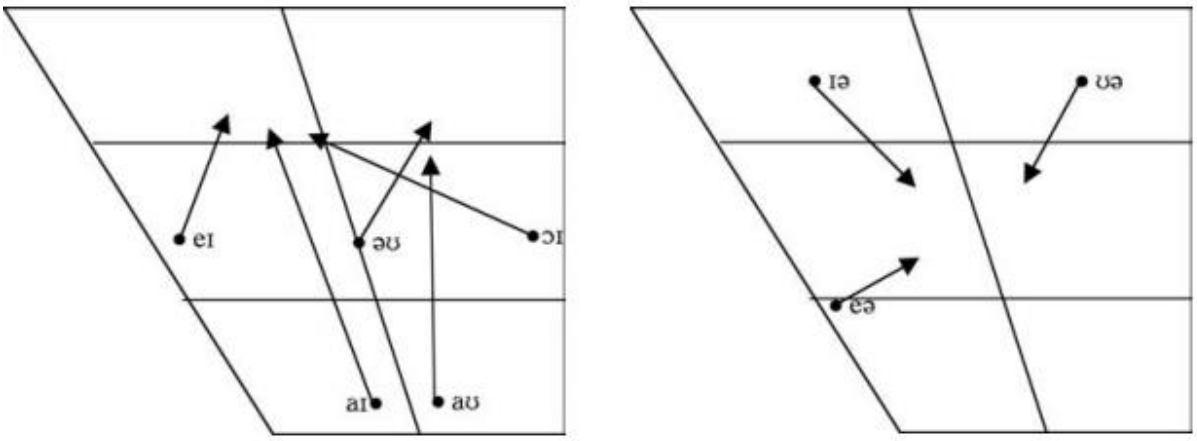


Fig. 2.9. *Closing and centring diphthongs in RP English (Roach 2004).*

2.4.3 Consonants

RP English employs 24 consonants distinguished based on place and means of articulation. There are pairs of consonants that share the place and means of articulation and differ only in being voiced or voiceless (e.g. /z/ *x* /s/). Voiceless plosives tend to be aspirated in stressed syllables except after /s/. "The sonorants /m, n, ŋ, l, r/ may be pronounced as syllabic consonants in place of a weak syllable containing a vowel, e.g. bottle/bɒtl̩/, button /bʌtn̩/" (Roach 2004: 241). Syllabic consonants are never present in Vietnamese (see 2.3.5.3).

	Bilabial	Labiodental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosive	p b			t d			k g	
Affricate					tʃ dʒ			
Nasal	m			n			ŋ	
Fricative		f v	θ ð	s z	ʃ ʒ			h
Approximant	(w)				r	j	w	
Lateral approximant				l				

Fig. 2.10. *Consonants in RP English (Roach 2004).*

2.4.3.1 Consonant clusters

It is quite characteristic for the English language to place more consonants together and create consonant clusters. The clusters can be placed initially as well as finally. Roach (2006) distinguishes two-consonant and three-consonant initial consonant clusters and two-, three- and four-consonant final clusters. Formation of consonant clusters in Vietnamese is impossible and therefore the Vietnamese speakers of English experience problems with pronunciation of such clusters.

2.4.4 RP English prosody

As English is not a tonal language, intonation (or pitch contour) is not employed in lexical distinction. Roach (2006) describes 5 tonal contours (level, fall, rise, fall-rise, rise-fall) with various pragmatic and emphatic functions. Such functions are often expressed by means of lexical elements in Vietnamese (see 2.3.8).

In RP English, "there is a very marked difference between weak, unstressed syllables which in some contexts may be almost undetectable and strong syllables (stressed or unstressed) which are fully pronounced" (Roach 2004: 243).

As opposed to the Vietnamese who appear to produce "chopped up" speech due to extensive use of glottal stops before initial vowels, native RP English speakers use certain linking phenomena helping to connect their speech. There are five means of linking (Volín 2005: 64-66): *a) pseudo-resyllabification* – seeming misalignment of syllabic and word boundaries occurring when word-initial vowels are preceded by word-final consonants; *b) linking [r]* - occurring when a word-initial vowel is preceded by silent 'r'; *c) intrusive [r]* – used to avoid intervocalic glottal stops and to prevent two vowels from a direct contact. Unlike *linking [r]* it is not represented in spelling; *d) transient [j]* – occurring when a word-initial vowel is preceded by word-final /i:/ or /ɪ/; *e) transient [w]* – occurring when a word-initial vowel is preceded by /u:/ or /ʊ/.

English rhythm is considered to be stress-timed, i.e. the intervals between stressed syllables tend to be alike and unstressed syllables are compressed to preserve the isochrony of the inter-stress intervals whereas in syllable-timed isochrony it should be preserved in inter-syllabic intervals.

2.5 Vietnamese English

As it is the aim of the empirical part of this thesis to describe the rhythmical features of the Vietnamese accent in English, this section will focus mainly on the segmental

features of Vietnamese English stated by Cunningham (2009) and Singer (2012). Morphological and syntactic issues will also not be addressed for space reasons and also because their relevance to pronunciation cannot be assessed as the recorded material consists of pre-prepared texts and not of impromptu speech. We will attempt to illustrate the features in question by examples taken from the material

The pronunciation of English presents severe challenges to Vietnamese-speaking learners. Not only is the sound system recorded for the purposes of our research on rhythm of Vietnamese very different from that of English, but there are also extremely limited opportunities for hearing and speaking English in Vietnam. In addition, there are limited resources available to teachers of English in Vietnam so teachers are likely to pass on their own English pronunciation to their students. University students of English are introduced to native-speaker models of English pronunciation, notably Southern educated British, but they do not often have the opportunity to speak with non-Vietnamese speakers of English (Cunningham 2009).

2.5.1 Vowels

The inventory of Vietnamese vowels is relatively rich and its speakers do not have difficulties adjusting the quality of English vowels but they tend to struggle with quantity because quantity is a distinguishing factor for only two pairs of Vietnamese vowels: */ă/; /ɤ̆/* x */a/; /ɤ/*. Occasionally, the long vowels */iː/; /uː/* tend to be pronounced like /ɪ/ and /ʊ/, hence the word *piece* sounds more like *piss* (JCAm2-07-02). The figure below describes yet another case of variation in vowel pronunciation.

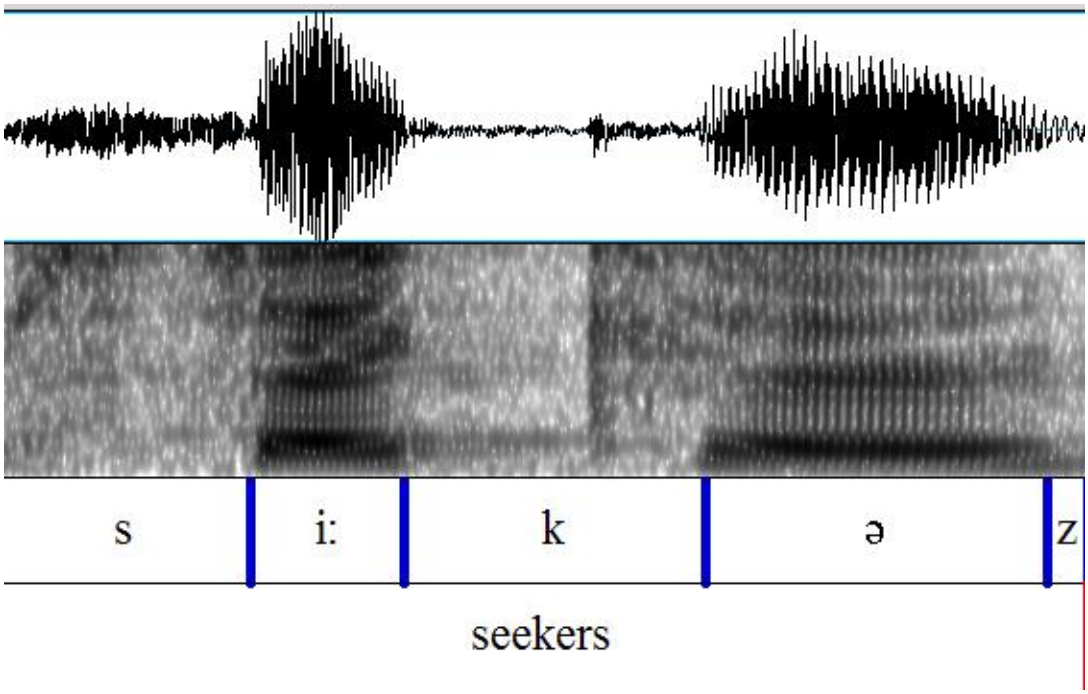


Fig. 2.11. *The long vowel /i:/ clearly has shorter duration than the weak vowel /ə/. Moreover, the pitch before /k/ is rising as the syllable contains the rising tone sắc. The spectrogram shows that the realisation of /i:/ is actually closer to /ɛ/ but that can be caused by the speaker's insufficient familiarity with the word. (AMAf1-04-01).*

2.5.2 Consonants

2.5.2.1 Simplification of consonant clusters

As a tonal monosyllabic language, Vietnamese does not employ consonant clusters in order to secure larger variability of lexemes. The lack of consonant clusters in their own language often renders the Vietnamese speakers incapable of pronouncing the clusters used in English, which leads to elision of some of the consonants especially plosives or fricatives towards the ends of syllables or words. Consonant clusters simplification might affect the rhythm of Vietnamese English. On the other hand, it seems in many cases that the missing consonants are realized by silence.

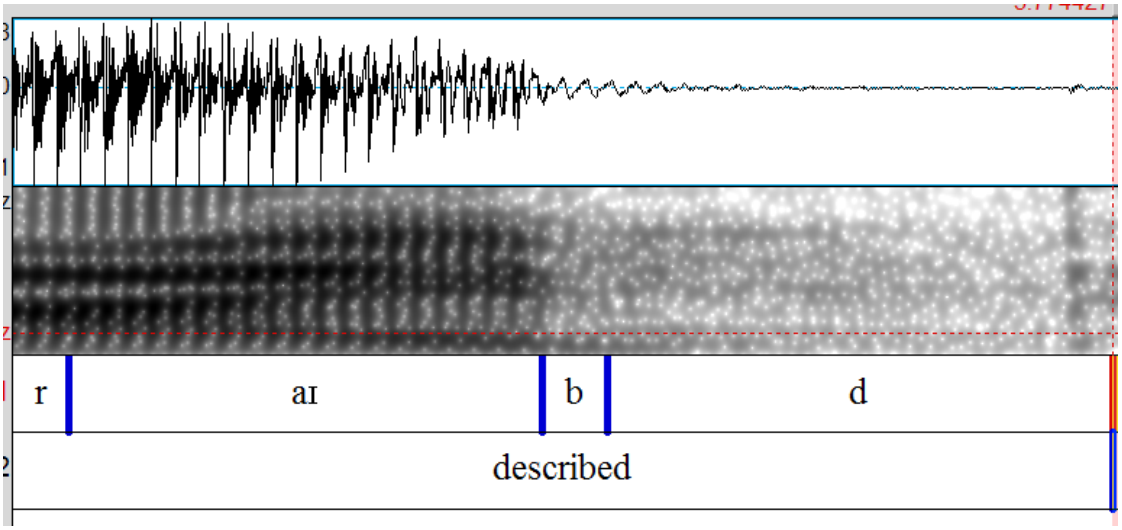


Fig. 2.11. *Elision of a lenis plosive /b/ and a very weak release of a lenis plosive /d/ due to linguistic interference from Vietnamese where there are no consonant clusters and the release of syllable-final plosives is almost inaudible and often implosive. (JAAm1-03-02).*

2.5.2.2 Final plosives

The semantically heaviest element in Vietnamese is the centre with the tonal vowel, ends of syllables have smaller semantic relevance. In English, on the other hand, there are many lexemes in which the semantic difference is revealed by their last consonant or consonant cluster. This poses a difficulty to many Vietnamese speakers who tend to "neglect" the pronunciation of syllable-final elements in English and hence reduce their intelligibility. "All final Vietnamese plosives [p, t, k] are glottally-reinforced [ʔpˀ, ʔtˀ, ʔkˀ]" (Singer 2012: 2), which causes the release to be very weak or even completely absent. Moreover, syllables ending with plosives can only carry the rising tone *sắc* or the glottalized falling tone *nặng*. These two features lead to very weak release of final plosives in English (or their possible complete omission) and to the fact that the vowel in the last syllable shows either a rising pitch or strong glottalization and sometimes noticeably falling pitch. The abrupt change of pitch (especially in case of the rising tone *sắc*) renders the vowel perceptionally shorter.

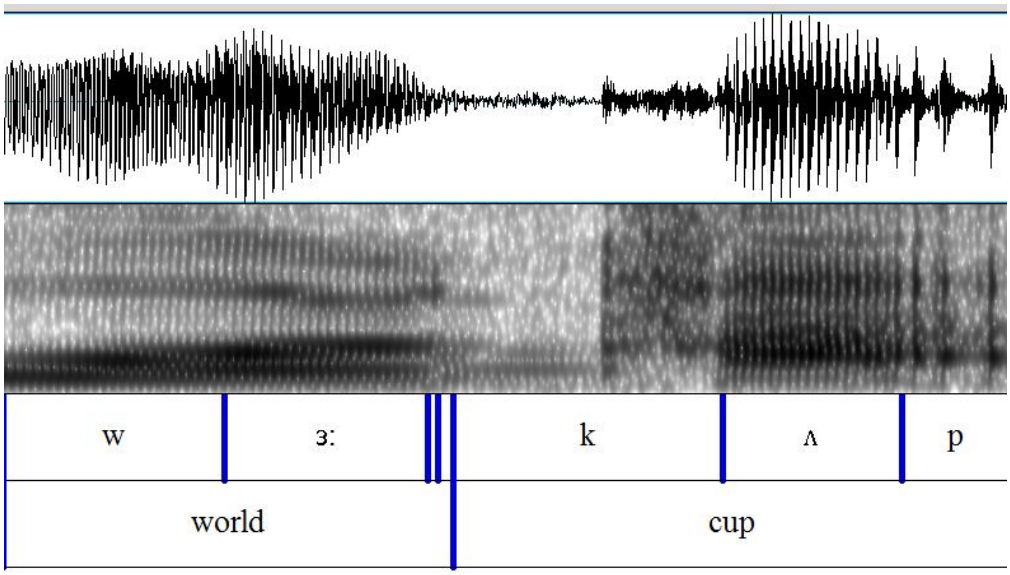


Fig. 2.12. *Complete elision of /l/ and /d/ in /wɜːld/ and heavily glottalized realization of /p/ in /kʌp/ (JAAf4-08-01).*

2.5.2.3 Initial Glottal Stops

"All vowel-initial words [in Vietnamese] begin with a glottal stop [ʔ], regardless of whether or not they are following a pause." (Singer 2012: 2) This feature is present in casual everyday conversation as well as in careful news announcements. As it was noted in chapter 2.4.4, the English language uses the phenomenon of linking rather extensively and so the speech is connected and void of word-initial glottal stops. Vietnamese speakers, however, usually place glottal stops in front of word-initial vowels and their speech therefore seems rather disconnected.

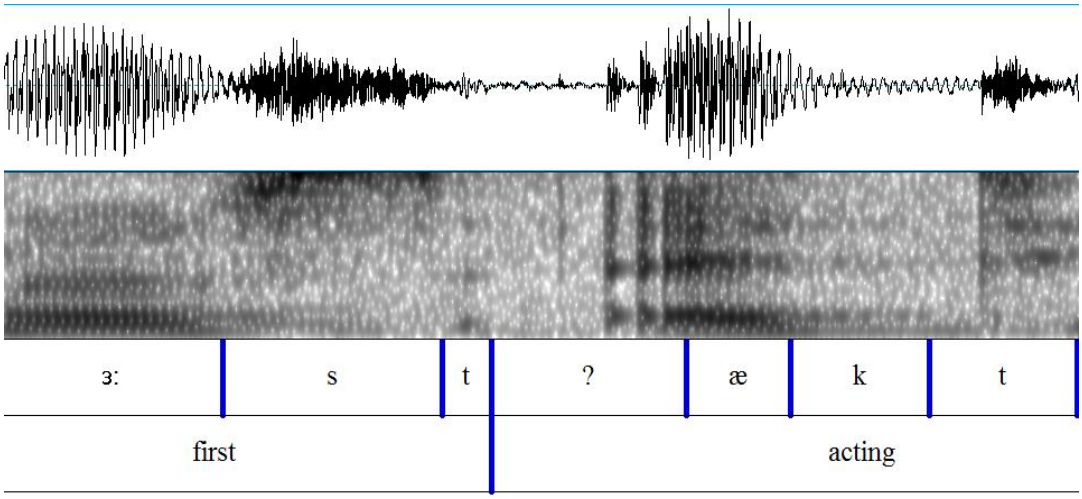


Fig. 2.13. *A glottal stop before an initial /æ/. A very weak release of final /t/ and no release of post-vocalic /k/ (JLAf3-02-02).*

2.5.2.4 Elision of Final [s] or [z]

Final [s] or [z] seem to be elided rather often (as there are no final voiced fricatives in Vietnamese, phonetic realization of the grapheme *s* is almost always [s]) but it can possibly be caused merely by relatively higher occurrence frequency compared to the other final consonants due to the fact that *-s* is a very popular grammatical marker. Therefore, although elision of [s] and [z] might be categorized as a subgroup of consonant cluster simplification, we decided to treat it separately due to the frequency and also due to its relation to 2.5.1.5 below where it seems that the Vietnamese speakers are aware of the fact that they drop a lot of these phonemes and so they attempt to make up for it by inserting [s] in places where it is redundant.

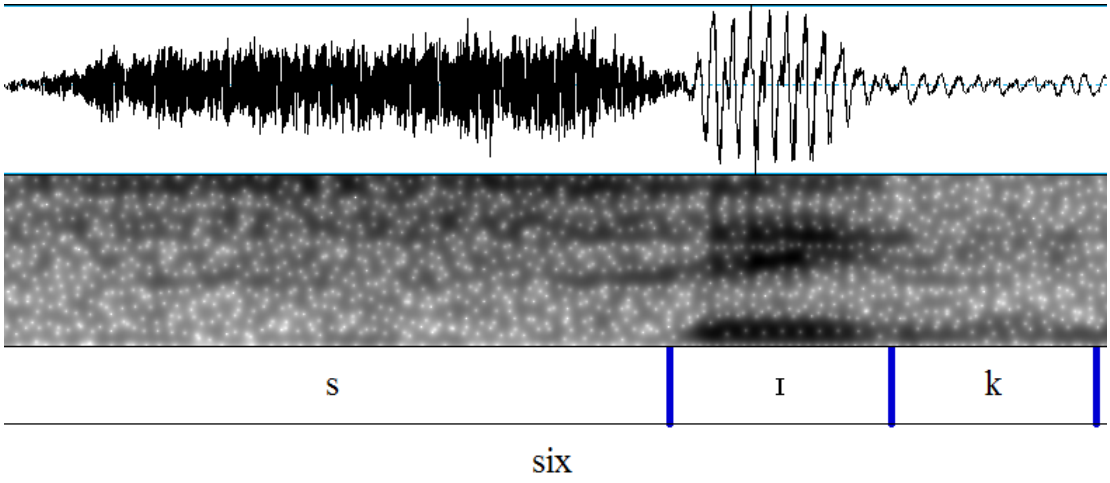


Fig. 2.14. *Elision of [s] after [k] in the word six. The original sentence* six people have been killed *therefore sounds more like* sick people have been killed *(JLAf2-06-02).*

2.5.2.5 Intrusive [s]

After analyzing all the material it turned out that the intrusive [s] was a recurring phenomenon throughout the speech of most of the recorded speakers. It would need further research in order to establish some kind of pattern but superficially it seems that we might be dealing with a case of hypercorrection.

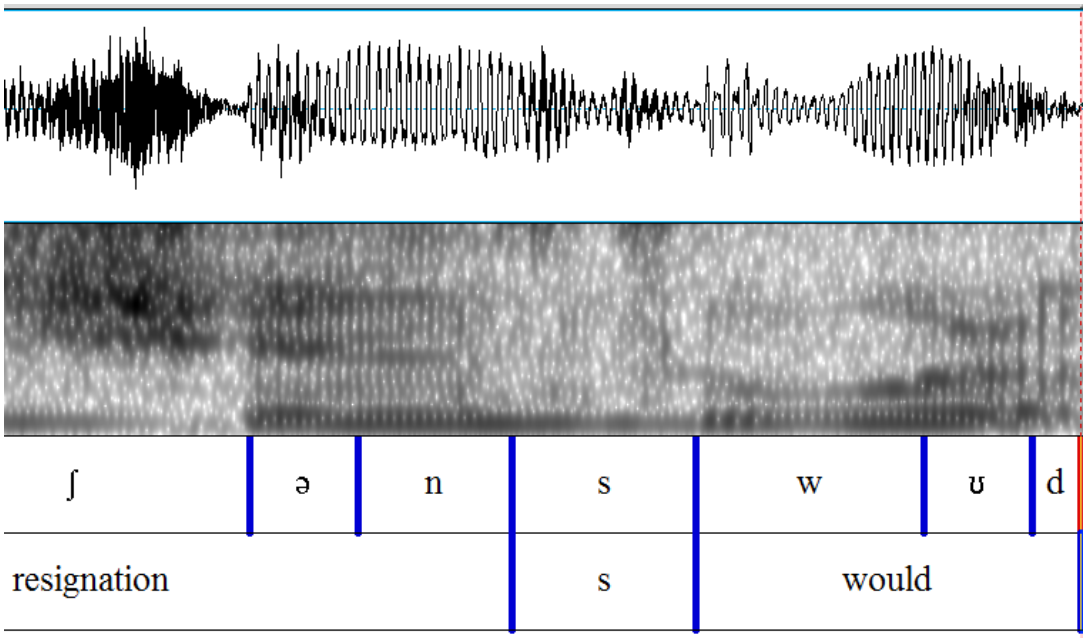


Fig. 2.15. *Intrusive [s] between the words resignation and would possibly caused by hypercorrection (AMAf4-09-02).*

2.5.2.6 /l/ x /n/ distinction

As it was noted in 2.3.5.3, speakers of Northern Vietnamese are prone to being unable to distinguish /l/ and /n/ in the syllable-initial position. This speech impediment then also hinders their acquisition of English as a foreign language. They struggle with pronunciation of words such as *need* x *lead* but also with words where /l/ x /n/ are constituents of consonant clusters such as *snack* x *slack*. It has been already mentioned that this problem is closely related to education and therefore there are no instances of it captured in our material as all of the respondents were highly educated especially considering Vietnamese standards.

3. Methodology

3.1 Speakers

Bearing in mind the secondary target of the project i.e. creating a sub-corpus of Vietnamese English, gathering a sufficient data pool was crucial. We managed to record 12 speakers for approximately 9 minutes each rendering close to 2 hours of material. The speakers were labelled by four capital letters. The first letter A-G stands for the order of speakers, after that VE stands for Vietnamese English and the last letter F/M represents gender. Speakers were ordered based on gender and age. For the complete list see table 3.1.

3.1.1 Selection of speakers

In selecting appropriate speakers it was necessary to consider the following features ranked according to importance:

3.1.1.1 Command of spoken English

Only speakers who were able to express themselves freely (at least B2 according to the Common European Framework of Reference for Languages) and whose annunciation was comprehensible were chosen for recording. After listening to some of the selected speakers it might seem that they are not very advanced. However, one must consider the European standards of assessing foreign language skills as the Asian standards are lower. Half of the speakers underwent long-term stays abroad. Two female speakers earned their BA titles at the Department of English, Hanoi State University. All of them had extensive experience in communication with foreigners and were active in the Hanoian international community. To draw a closer comparison, their level of English measured by the standards of the general Vietnamese public can be compared to the level of students at the Department of English or Translatology in regard to the general population in the Czech Republic.

3.1.1.2 Gender

The ratio of speakers is 7 females to 5 males so the gender representation is not markedly disproportional. It must be noted that persuading male speakers to participate in similar projects that do not bring any monetary reward tends to be more difficult than in the case of female speakers. On the other hand, female speakers exhibit more shyness and nervousness, which makes recording more problematic.

3.1.1.3 Origin

The pronunciation of Vietnamese across dialects varies to a great extent on the segmental as well as suprasegmental level. Therefore, in order to prevent as many variables as possible from corrupting the data, most speakers were originally from Hanoi or close proximity. Only one speaker came from Central Vietnam (city of Đà Nẵng) but lived in Hanoi for more than 5 years.

3.1.1.4 Age

Having a homogenous age group is always an advantage in phonetic research because speech organs undergo changes with time that could potentially alter the results. In our case, the homogenous age group is a consequence of the fact that the access to younger generations was limited and people 35+ either do not speak English or they are busy making money using their language skills.

CODE	REC. TAG	GENDER	AGE	ORIGIN	EXPOSURE SPECIFICATION
AVEF	AMAf3; JCAf1	F	23	Nam Định	B.A. in marketing; never left VN
BVEF	AMAf2; JAAf1	F	24	Hà Nội	B.A. in business, never left Vietnam
CVEF	AMAf4; JAAf3	F	24	Hà Nội	B.A. in English, teacher at elementary school, never left VN
DVEF	AMAf1; JLAf1	F	25	Đà Nẵng	B.A. in English , translator, never left Vietnam
EVEF	JAAf4; JLAf4	F	26	Hải Dương	B.A. in international trade; never left VN
FVEF	JAAf2; JLAf2	F	27	Hà Nội	M.A. from the UK and France; 2 years
GVEF	JCAf2; JLAf3	F	30	Hải Dương	B.A. in international trade; never left VN
AVEM	JCAm4; JLAm3	M	24	Hà Nội	2 years in the US; B.A. in business
BVEM	JAAm1; JLAm2	M	27	Hà Nội	M.A. from Netherlands; 2 years
CVEM	AMAm1; JCAm2	M	27	Thái Bình	M.A. from Netherlands, 2 years
DVEM	JCAm1; JLAm1	M	28	Hà Nội	Studied at ČVUT, CZ; Hanoi international community organizer
EVEM	JAAm2; JCA m3	M	32	Thanh Hoá	4 years in Australia, IT specialist

Tab. 3.1. *Table of recorded speakers listing their tags, gender, age, place of origin and the nature of exposure to English.*

3.2 Material

The speakers were asked to read BBC news (each speaker about 1000 words) from the database that members of the Department of Phonetics in Prague frequently use for both academic and pedagogical purposes. Namely, the texts labelled AMA, JLA, JCA and JAA were chosen (AMA-Alice Moss, JLA-Jackie Leonard, JAA-Jill Anderson, JCA-July Candler). Each text comprised approximately 500 words. The

texts were chosen so that the gathered data could be contributed to the growing phonetic corpus based on these BBC texts.

3.2.1 Material recording

All recordings were captured using the device MEDELI DR2, a device similar to Edirol R-09HR available for use at the Department of Phonetics. MEDELI DR2 seems to be a reliable field recording device. It has 4 microphones allowing stereo recording, sampling frequency up to 48 000 Hz, MP3 (up to 640kb)/WAV(16/32bit) format and a button to adjust sound input. In comparison with Edirol, it can record significantly larger amounts of data within one battery cycle and its price is only about 25%. The difference in recording quality shall be left to experts to assess but the recordings are acceptable for this kind of analysis. The format WAV 16 bit and sampling frequency 48 000 Hz was set as default setting for all the recordings.

My dormitory room served as the recording studio. It was a rather emergency solution but I tried my best to accommodate the room to fit the basic requirements of sound recording. A bed sheet was spread on the table used for recording. The recorder was installed into a small tripod and tipped slightly, so the microphones would face the speaker, then placed on the table. Pillows were placed around the recorder and a large towel was hung against the wall in order to prevent the sound waves from bouncing off the walls and creating echoes. The recorder was placed roughly 30 centimetres from the pillows and 50 centimetres from the towel. Space for the sheet with the text was created under the microphone in order to prevent rustling.

Speakers were seated approximately 50 centimetres in front of the recorder and they were asked to read the first part of the document as a test of sound intensity. The sound input was then adjusted accordingly. Unfortunately, even after the adjustment the intensity of the speakers differed to a great extent. As a rule, male speakers tended to speak more and more loudly during the recording. Female speakers displayed the opposite tendency, which was probably conditioned by Vietnamese cultural stereotypes where women are timid, quiet and humble until they get married. None of the respondents was married.

The speakers were given the texts to read silently in order to get acquainted with them. Then we instructed the speakers to read as clearly as possible at their natural pace. At first, they were also asked to reread sentences in which they made mistakes but such approach was abandoned due to infeasibility. Most speakers were not fluent enough to read through the text flawlessly; moreover, the text presented a number of exotic names whose pronunciation can be ambiguous even to native speakers. Some of the speakers (mostly female) suffered from stress induced by stage fright, being asked to read out loud in a non-native language in front of a male foreigner, and the indoor temperature reaching over 30 degrees Celsius.

There were few unexpected technical difficulties that must, unfortunately, be taken in consideration: a) the room was right next to a basketball court, the window was far from soundproof and so the sound of the ball hitting the basket or dribbling can be heard occasionally; b) Hanoi is generally very noisy due to the exorbitant number of motorcycles, engines create low-pitch noise that the microphones do not pick up but honking of a motorcycle as a high-pitch noise can appear; c) the room was equipped with a ceiling fan that kept the indoor temperature on a bearable level but it had to be switched off to avoid sound pollution of the recorded material, which rendered the atmosphere in the room rather discomforting and might have influenced the performance of speakers; e) by choosing my own dormitory room (in a building for foreign students) as the venue of recording I deprived myself of several female speakers whose upbringing and cultural background did not allow them to enter a room of a male foreigner alone.

Despite the fact that there is a Department of Phonetics at the University of Social Sciences and Humanities in Hanoi, applying for a permission to use their recording lab (the state and equipment of which is doubtful) was out of the question for the administrative reasons. Furthermore, it would have been extremely tedious to organize all speakers to come to the lab at a certain hour.

3.2.2 Material processing

The recordings were transferred from the device on a computer and labelled according to the abbreviations of BBC presenters and gender of the speaker. As each speaker was asked to read two presenters, we needed to note down which speaker read which text to avoid subsequent identification difficulties.

3.2.2.1 Sound Forge

As the second step we used Sony Sound Forge Audio Studio 8.0 to convert the data from stereo to mono as the stereo signal was too voluminous and, more importantly, redundant for the purposes of our analysis. In each case we selected the channel that presented a better quality signal. The sampling frequency was lowered to 32 000 Hz with the anti-alias filter applied.

As it was mentioned above, we experienced difficulties with signal intensity so we had to normalize it i.e. boost the weak signals and decrease the strong ones. The strong signal was in some places so intensive that it exceeded the caption capacity of the recorder and so the heights were lost beyond remedy. Fortunately, these places were relatively scarce.

Finally, we performed first rough cuts of the signal in order to process it further in PRAAT (Boersma and Weenink, 2011) because processing of larger files in PRAAT proves to be uncomfortable. These cuts were made in accordance with the paragraphs in the BBC texts.

3.2.2.2 Breath Group segmentation in PRAAT

The paragraphs were further subdivided into smaller units labelled BG (breath groups). Ideally, these units should capture the speech signal between the speaker's inhalation. However, given that the speakers were neither trained orators nor native speakers, the division was mostly based on syntactic structure merely in order to preserve semantic coherence of the text and maintain a reasonable length of the

material. The inhalation patterns of most recorded speakers were rather unpredictable and unnatural. The length of the individual breath groups spans from 2 to 16 seconds.

3.2.2.3 Forced alignment

The breath groups were then run through the Penn Phonetics Lab Forced Aligner (Yuan-Liberman 2008) that was supposed to establish the boundaries between the individual speech sounds. After applying the aligner, we ended up with textgrids containing 4 tiers: *breath group*, *word*, *stress* (the stress marks were placed automatically based on dictionary definitions and they represented the Vietnamese stress patterns only very little; fortunately, this tier was not crucial for our analysis) and *phone*.

However, a large amount of time (approx. 200 hours) had to be invested in the manual segmentation mainly due to the large number of dysfluencies and hesitations that the aligner was not able to deal with. Another hindrance preventing the programme from more accurate work was the exotic names that were not included in the programme's dictionary, abbreviations and numerals (at least the problem with numerals could have been avoided by simply transcribing them in words). The programme skipped all these items and labelled them "sp" (sound pause), which meant that their segmentation had to be carried out fully manually. The aligner was also programmed to capitalize all words that exhibited signs of dubious segmentation. As the material was not American English and, moreover, there were numerous instances of dysfluencies and hesitations, almost all the words showed in PRAAT as capitalized. This could have been remedied by a script that would change all the capitalized letters into lower case but after adopting a foolish decision to change everything manually, approximately 10 hours or work was lost in total.

3.2.2.4 Manual segmentation

As it was touched upon in the previous section, the forced aligner considered a tremendous help in segmentation of the material but as it was "trained" for fluent American English, a significant amount of manual correction was necessary. The correction was carried out in the following steps:

a) Getting rid of capital letters and establishing word boundaries – In this step we replaced capital letters with lower case and erased all labels "sp" (speech pause). Subsequently, we re-labelled the empty segments *{pause}* whenever there was audible inhalation and/or the pause was longer than 400 milliseconds, *{hes}* when there was a rhythmical disfluency in form of silence (longer than 120 milliseconds) or *{dsfl}* whenever there were incomplete words or extra speech sounds. When the empty segment contained speech sounds of a neighbouring word, they were added to the respective word. Whenever the empty segment contained silence shorter than 120 milliseconds, it was either labelled as a glottal stop (if preceded by a vowel) or it was divided in half and added to the preceding and following segments.

b) With the help of *Longman Pronunciation Dictionary* (Wells: 2003), the spelling suggested by the aligner was transformed into the British standard including the rules concerning vowel reduction in connected speech. The recorded material obviously does not entirely correspond to the transcription. There are differences mainly in vowel quality and place of articulation of fricatives. However, this thesis focuses on rhythm and not on the particular realization of speech sounds hence the V-C durations were the main concern of the segmentation.

c) Speech sound boundaries – As the last step we established the boundaries of the individual speech sounds along the guidelines suggested by Machač and Skarnitzl (2009). The words were segmented according to the dictionary and the rules for vowel reduction in connected speech. Probably as interference from Vietnamese there were problems with linking the words together, which resulted in the presence of many glottal stops in front of vowels. Although the British standard is non-rhotic, if the Vietnamese speaker was rhotic, it was marked in order not to corrupt the rhythmical pattern. The same applies to minor pronunciation errors such as the word "bombing" being pronounced as [bɒmbɪŋ]. Major pronunciation errors were

excluded from the analysis by adding the tag *{alt}* to the segment in the tier *word*. After having added the tag *{alt}*, it was crucial to remove all the segments from the topmost tier labelled *phone* in order for the calculations of the rhythm not to be compromised.

3.2.2.5 Preparation for analysis

Before calculating and the parameters %V; ΔV; ΔC; varcoV; varcoC; rPVI-V; rPVI-C; nPVI-V and nPVI-C it was necessary to prepare a table in which the results could be assorted and evaluated. We created a MS Excel file and copied the recording tags (e.g. *AMAf1-01-01*) into the first column. As we had recorded 24 files with merely 12 speakers, each speaker had read 2 files and hence had been assigned two recording tags. For this reason it was necessary to add an extra column *Speaker* and fill in labels according to the table in 3.1.1 (e.g. *AVEF*, *EVEF*). The third column was used for gender distinction *f/m*. The fourth column served to distinguish length and was dependent on the column *duration*. If the duration was less than 2 seconds, the segment was labelled *short*. If it was equal to or longer than 2 seconds, the segment was labelled *long*. The fifth column was labelled *prosody* and it was devised to assess prosodic compactness of the speaker on a scale from 0 to 3 (least compact to most compact). There were two decisive criteria for the assessment: number of word-initial glottal stops in each utterance and the subjective judgement after careful listening of the material. After that, we also assessed the proportion of dysfluencies with the individual speakers on the scale from 1 (largest proportion) to 5 (smallest proportion). The proportion of dysfluencies (segments in the tier *word* labelled *{pause}*, *{hes}* and *{dsfl}*) within each BG was calculated in Excel. As the last step we labelled additional columns according to the rhythm metrics mentioned above.

Breath group *JAA-f4-10-01* was excluded from the analysis because the pronunciation was very careless and the results were not comparable with the rest of the data.

3.2.2.6 Extracting data

In order to extract the values for duration of vocalic and consonantal intervals, we used a series of PRAAT scripts created by members of the Department of Phonetics and Phonology at the Charles University in Prague.

Since we were interested in duration of vocalic and consonantal intervals, the actual realizations of the individual phonemes was irrelevant and hence the first script was used to convert patterns such as CCCVVC to CVC.

Different scripts were used to extract the figures for nPVI-c, nPVI-v, rPVI-c, and rPVI-v (according to Low and Grabe) and %V, ΔV, ΔC (according to Ramus et al.) The figures for varcoV and varcoC were calculated subsequently (see 2.2.3).

3.3 Hypotheses

H_{A0} – Vietnamese English displays similar values for all rhythm metrics as RP English.

H_{A1} – Vietnamese English displays higher consonantal values due to the absence of linking and frequent glottal stops preceding word-initial vowels.

H_{B0} – 5 of the recorded speakers were male and 7 were female; however, Vietnamese English does not manifest any differences in rhythm between genders

H_{B1} – we might find patterns indicating differences in rhythm between genders.

H_{C0} – frequency and duration of dysfluencies have no effect on rhythm in Vietnamese English

H_{C1} – frequency and duration of dysfluencies influence the rhythm in Vietnamese English.

H_{D0} – the rhythm measurements do not vary according to the degree of compactness of the speakers.

H_{D1} - the results should vary according to the degree of compactness. The more compact the speaker is the more the figures should resemble the figures measured with the RP speakers.

H_{E0} - Vietnamese English and the RP standard do not differ in terms of the stress-timed *x* syllable-timed dichotomy.

H_{E1} - based on Cunningham's (2009) claim that Vietnamese seems to be a language with a tendency to syllable-timing, the Vietnamese speakers of English are expected to produce values indicating a stronger tendency towards syllable-timing than the RP speakers

4. Analysis

In this chapter we will present the results computed from the data we retrieved from the recordings. Firstly, we will compare arithmetic means of all the values described in section 2.2 with the values estimated for British standard. Secondly, we will present the individual categories (%V, ΔC, ΔV etc.) organized according to the variables: *speaker*, *gender*, *dysfluency* and *compactness*. In order to ensure the reliability of results, we will perform two separate average calculations for each variable. The first calculation will contain all breath groups whereas the second will exclude all breath groups shorter than 2 seconds. By doing so, the results should be rendered more accurate because short breath groups tend to be inconsistent and can corrupt the results.

Obtaining the referential data describing the British standard proved to be rather complicated. After comparing the data presented in various studies on rhythm (Ramus et al. 1999, Grabe&Low 2002, Wagner&Delwo 2004, White&Mattis 2007, Mairano&Romano 2011) we realized that the data differed from each other to such an extent that they were no longer comparable to our results. The significant discrepancy in the data across the studies might have been caused by different means of calculation. The studies never mention any concrete procedures of calculation and there is therefore no way to check whether we followed exactly the same method. Another reason for this discrepancy might be seen in the amount of recorded material in each study. Every speaker possesses their own idiosyncratic means of communication might not be ultimately comparable to other speakers. Therefore, the figures mentioned in the studies do not necessarily represent British English as a whole but only a sample of a small number of speakers.

In order to gain data that would be comparable to the data we gathered from the Vietnamese English, we used the same methodology on the original recordings of BBC news readers and we extracted the referential data from there using PRAAT and PRAAT scripts as described in Chapters 3.2.2.3 – 3.2.2.6.

4.1 Vietnamese English versus British Standard

We firstly present the data gathered from the Vietnamese speakers (including standard deviation) divided into consonantal and vocalic measurements. Most studies on rhythm copy the original methodology of Grabe and Low's discarding the figures rPVI-V and nPVI-C and comparing only rPVI-C and nPVI-V. We will not conform to this approach and the figures for rPVI-V and nPVI-C for Vietnamese English will also be taken in consideration.

%V	$Sd_{\%V}$	ΔV	$Sd_{\Delta V}$	varcoV	Sd_{varcoV}	rPVI-V	Sd_{rPVI-V}	nPVI-V	Sd_{nPVI-V}
38,8	**4,5**	**52,7**	**16,8**	**55,7**	**12,5**	**56,4**	**18,1**	**56,8**	**11,8**

Tab. 4.1. *Vocalic measurements of Vietnamese speakers of English including standard deviation.*

ΔC	$Sd_{\Delta C}$	varcoC	Sd_{varcoC}	rPVI-C	Sd_{rPVI-C}	nPVI-C	Sd_{nPVI-C}
75,4	**17,4**	**57,0**	**8,8**	**80,6**	**20,1**	**60,4**	**10,4**

Tab. 4.2. *Consonantal measurements of Vietnamese speakers of English including standard deviation.*

The consonantal and vocalic measurements presented in tabs 4.1. and 4.2. were subsequently compared to the data from the original BBC recordings processed at the Department of Phonetics in Prague and a graph was created for better orientation.

	%V	**ΔV**	**ΔC**	**varcoV**	**varcoC**	**rPVI-V**	**rPVI-C**	**nPVI-V**	**nPVI-C**
VE	38,8	52,7	75,4	55,7	57,0	56,4	80,6	56,8	60,4
BrE	40,3	47,2	59,7	59	53,1	50	66,8	61,2	58,9

Tab. 4.3. *Differences in rhythm measurement parameters between Vietnamese English and British English.*

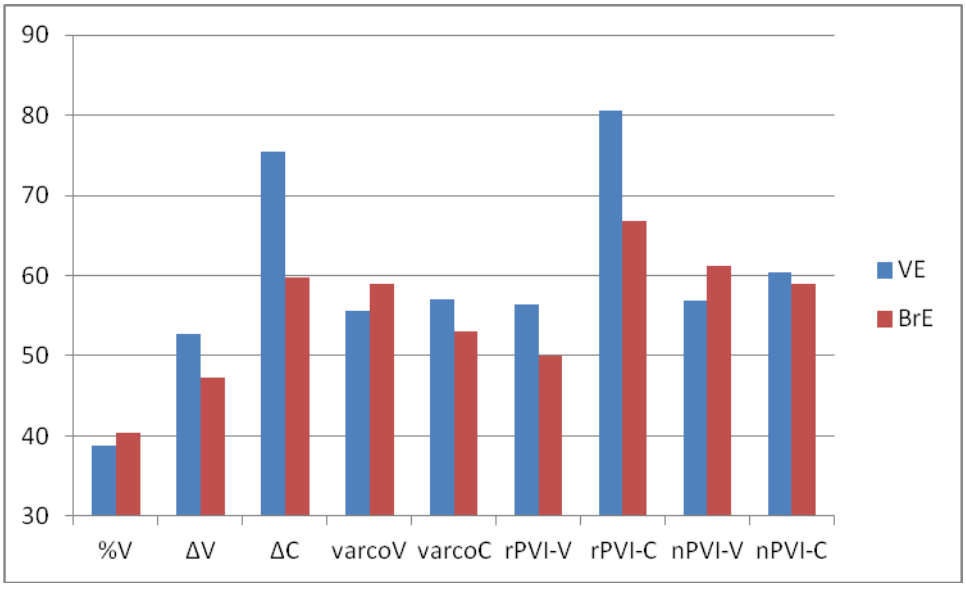


Fig. 4.1. *Rhythm measurement differences between Vienamese English and British English.*

In the figure above we can notice that vocalic measurements that are considered relevant by most studies on rhythm (i.e. %V; varcoV and nPVI-V) are consistently lower for Vietnamese English than for RP. On the other hand, relevant consonantal measurements (ΔC; varcoC and rPVI-C) are higher for Vietnamese English than for RP. The difference in consonantal measurements is greater than the difference in vocalic measurements. Especially ΔC and rPVI-C display visually prominent differences.

4.2 Speakers

Now we will present and compare the average results by individual speakers.

Speakers	%V	ΔV	ΔC	varcoV	varcoC	rPVI-V	rPVI-C	nPVI-V	nPVI-C
AVEF	42,2	54,5	68,9	54,1	55,8	57,8	74,4	55,1	56,2
AVEM	38,1	50,5	72,4	52,5	51,7	51,2	77,1	50,1	63,6
BVEF	39,6	48,4	69,8	55,1	57,6	50,8	70,2	56,0	61,4
BVEM	37,8	62,3	85,8	59,8	57,9	66,8	94,8	58,8	57,8
CVEF	38,8	58,9	78,3	60,6	58,2	65,2	83,6	63,9	63,6
CVEM	37,2	49,3	70,1	59,1	59,2	52,2	75,2	58,4	62,6
DVEF	40,8	44,0	65,0	48,5	54,9	47,9	69,1	52,8	55,0
DVEM	34,4	38,2	75,6	49,9	59,1	41,4	81,5	50,9	59,3
EVEF	41,1	60,6	89,2	55,2	61,4	66,2	95,2	57,2	59,5
EVEM	40,5	64,5	76,9	64,7	56,3	69,5	81,2	67,5	59,7
FVEF	38,1	54,7	81,5	54,3	55,4	55,8	82,4	53,0	62,9
GVEF	37,7	49,0	71,0	56,1	56,1	52,9	79,2	58,0	60,5

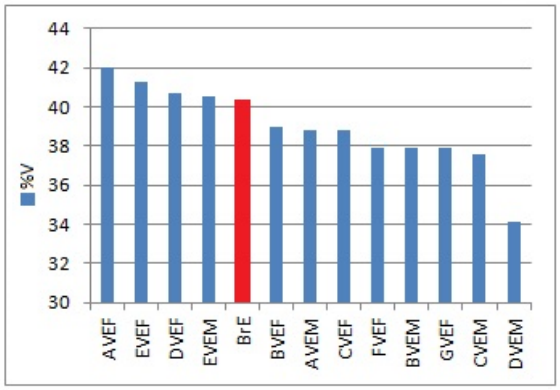
Tab. 4.4. *Averages of rhythm metrics by individual speakers including short BGs.*

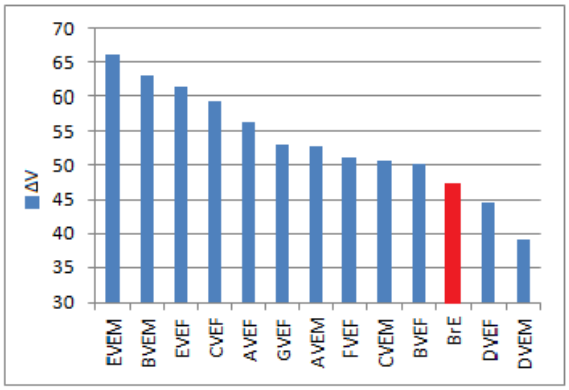
Speakers	%V	ΔV	ΔC	varcoV	varcoC	rPVI-V	rPVI-C	nPVI-V	nPVI-C
BrE	*40,3*	*47,2*	*59,7*	*59*	*53,1*	*50*	*66,8*	*61,2*	*58,9*
DVEF	40,7	44,6	65,4	49,4	55,0	48,8	70,0	54,1	59,6
CVEM	37,6	50,8	69,2	60,7	59,1	53,0	75,2	59,0	63,4
AVEF	42,0	56,4	70,0	56,1	55,5	58,9	76,8	55,9	59,9
BVEF	39,0	50,3	71,6	57,6	57,7	53,1	74,6	58,9	59,5
GVEF	37,9	53,0	71,9	60,1	56,1	56,8	82,2	61,1	62,9
AVEM	38,8	52,9	72,8	54,6	52,1	53,6	79,1	51,5	56,0
DVEM	34,1	39,2	76,4	51,7	59,6	41,9	83,0	51,9	62,1
CVEF	38,8	59,4	77,9	61,1	58,0	65,7	83,9	64,2	63,9
EVEM	40,6	66,2	78,3	65,8	56,7	71,7	84,4	68,8	61,9
FVEF	37,9	51,8	80,2	52,8	55,4	54,6	85,3	53,7	59,1
BVEM	37,9	63,2	86,9	60,3	58,4	67,9	96,9	59,4	63,5
EVEF	41,3	61,3	89,4	55,5	61,6	67,1	96,2	57,3	64,1

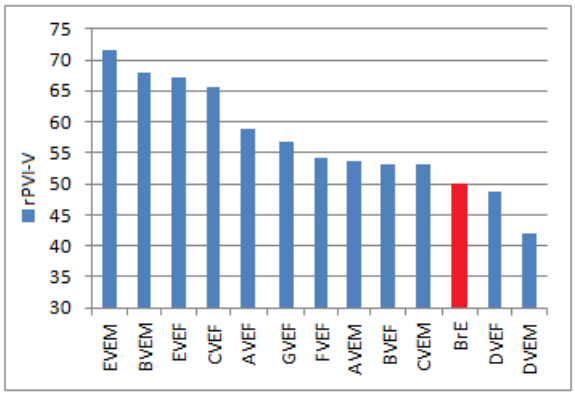
Tab. 4.5. *Averages of rhythm metrics by individual speakers with BGs shorter than 2 sec. filtered out. The table also includes the figures for the referential British RP standard.*

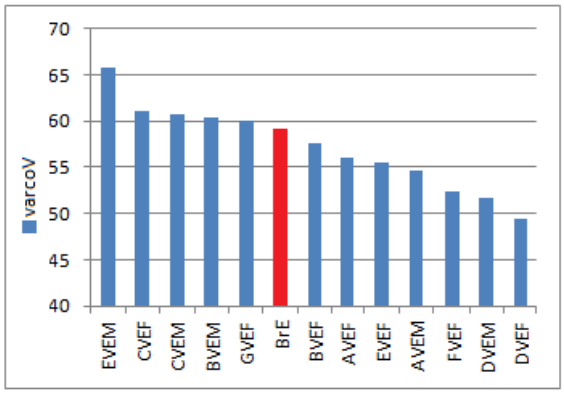
The tabs show higher values of consonantal measurements in case of Vietnamese English. Especially the speaker EVEF displays markedly higher values. The differences of vocalic measurements are not as prominent.

Graphs and ANOVA analyses of vocalic and consonantal measurements are presented below.









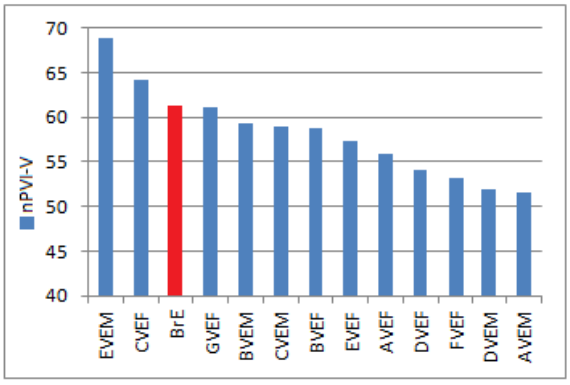


Fig. 4.2. *Graphic portrayal of vocalic measurements for individual speakers Vietnamese English (blue) based on tab. 4.5. and the average of the referential BrE RP accent (red).*

Parameter	ANOVA	Highest deviation (Tukey's test)
%V	$F(12, 522) = 13.43$; $p < 0{,}001$	DVEM; AVEF; EVEF
ΔV	$F(12, 522) = 13.36$; $p < 0{,}001$	DVEM; DVEF; EVEM
varcoV	$F(12, 522) = 9.01$; $p < 0{,}001$	DVEF; EVEM
rPVI-V	$F(12, 522) = 14.7$; $p < 0{,}001$	DVEM; EVEM; BVEM
nPVI-V	$F(12, 522) = 12.94$; $p < 0{,}001$	EVEM; CVEF

Tab. 4.6. *Analysis of variance carried out on the vocalic measurements. Speakers expressing the highest degree of deviation are listed in the right column.*

The graphs above suggest that the difference between vocalic measurements for VE and BrE is not as prominent as for consonantal measurements (see below). Such assumption is supported by the differences in values of BrE and the individuals with the highest degree of deviation. Whereas the differences in vocalic measurements oscillate around 15 units, differences in consonantal measurements reach up to 30 units. The two parameters described by most researchers as most influential in terms of vocalic measurements of rhythm (%V and nPVI-V) indicate that a substantial part of Vietnamese speakers of English achieve similar vocalic patterns as the speakers of RP. The ANOVA test suggests that speakers EVEM and DVEM express the highest degree of deviation. EVEM deviates in the direction of higher values whereas DVEM deviates in the direction of the lower numbers.

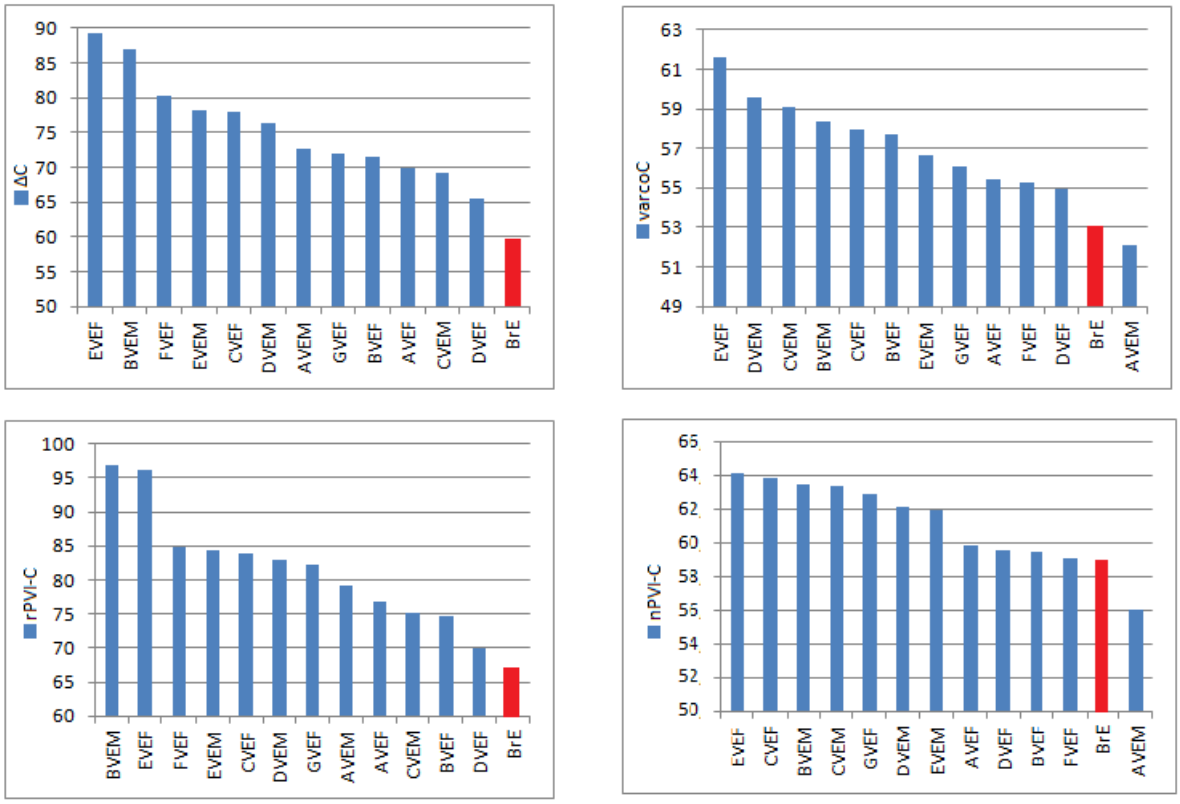


Fig. 4.3. *Graphic portrayal of consonantal measurements for individual speakers Vietnamese English (blue) based on tab. 4.5. and the average of the referential BrE RP accent (red).*

Parameter	ANOVA	Highest deviation (Tukey's test)
ΔC	$F(12, 522) = 8.61$; $p < 0{,}001$	EVEF; BVEM
varcoC	$F(12, 522) = 3.59$; $p < 0{,}001$	AVEM
rPVI-C	$F(12, 522) = 9.59$; $p < 0{,}001$	DVEF; EVEF; BVEM
nPVI-C	$F(12, 522) = 3.2$; $p < 0{,}001$	AVEM

Tab. 4.7. *Analysis of variance carried out on the consonantal measurements. Speakers expressing the highest degree of deviation are listed in the right column.*

Figure 4.3. above introduces the tendency of higher values of consonantal measurements with Vietnamese English speakers. Especially in the case of ΔC and

rPVI-C it is apparent that Vietnamese speakers positively deviate from the values of the RP standard. According to ANOVA, speakers EVEF and BVEM display values that deviate even from the average Vietnamese English values. Speaker DVEF, on the other hand, displays values rather similar to the RP standard.

4.3 Gender

The tables and graph below represent rhythmical differences based on gender. Languages can display differences in speech parameters between genders and therefore we decided to investigate this issue in regard to Vietnamese English.

Gender	%V	ΔV	ΔC	varcoV	varcoC	rPVI-V	rPVI-C	nPVI-V	nPVI-C
f	39,7	53,8	75,3	57,3	57,6	57,7	79,7	57,2	60,4
m	37,9	51,7	75,5	56,6	55,2	55,2	81,5	56,3	60,5

Tab. 4.8. *Table of average rhythm measurements for male and female speakers of Vietnamese English including short BGs.*

Gender	%V	ΔV	ΔC	varcoV	varcoC	rPVI-V	rPVI-C	nPVI-V	nPVI-C
f	39,7	55,5	76,3	57,3	57,3	59,5	82,1	58,7	61,2
m	38,1	53,2	76,3	57,1	57,0	56,7	83,5	57,3	61,5

Tab. 4.9. *Table of average rhythm measurements for male and female speakers of Vietnamese English with short BGs filtered out.*

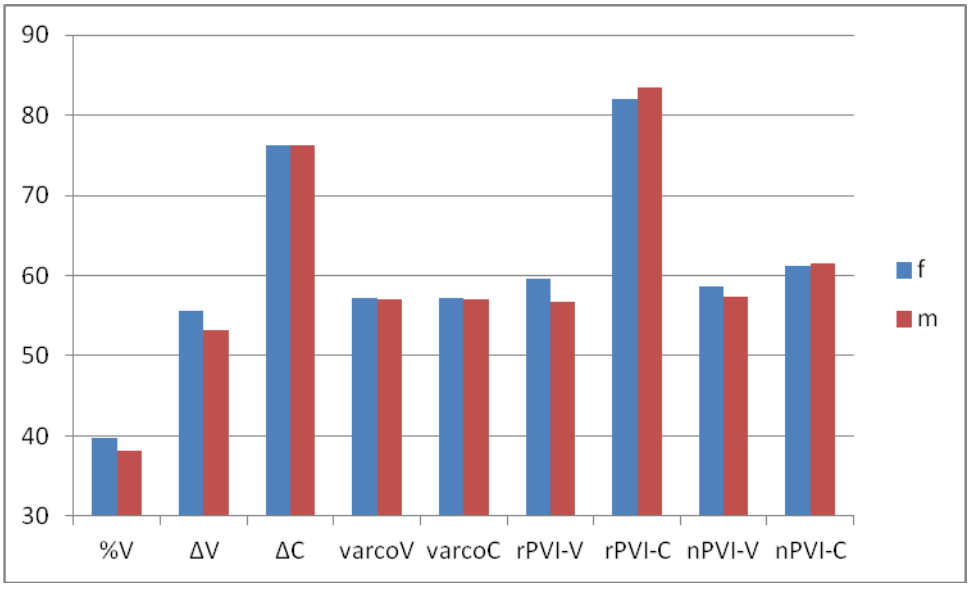


Fig. 4.4. *Graph of average rhythm measurements for male and female speakers of Vietnamese English based on tab. 4.9.*

The graph shows that vocalic measurements tend to be higher for female speakers and consonantal measurements appear slightly higher for male speakers. However, according to the ANOVA calculations, only the differences in %V carry statistical significance: $F(1, 533) = 20.74$; $p < 0{,}001$. The rest of the parameters are statistically insignificant ($p > 0{,}05$).

4.4 Dysfluency

This category was assessed based on the proportion of dysfluency duration to the overall duration of each individual breath group. After processing the data there turned out to be no noteworthy correlation, which might be explained by the fact that the dysfluencies constituted reading errors caused by the unfamiliarity of the speakers with the texts. The dysfluencies were mainly present in regional and proper names and their knowledge cannot be taken for an indicator of the speaker's skill. The dysfluencies were calculated in terms of duration but they were excluded from the analysis completely.

4.5 Compactness

The table and graphs below present the results based on prosodic compactness of the individual speakers, i.e. logical and natural breathing, linking and frequency of glottal stops (see 3.2.2.5)

Comp.	%V	ΔV	ΔC	varcoV	varcoC	rPVI-V	rPVI-C	nPVI-V	nPVI-C
BrE	40,3	47,2	59,7	59	53,1	50	66,8	61,2	58,9
three	39,8	44,6	69,5	48,0	55,9	48,4	70,1	50,4	56,3
two	39,1	54,0	75,9	56,3	56,7	57,8	81,6	57,6	60,4
one	38,1	54,1	76,8	57,6	57,4	57,7	83,2	58,2	61,8
zero	37,8	55,0	77,6	58,5	59,1	58,2	83,8	58,1	63,4

Tab. 4.10. *Mean rhythm metrics split according to prosodic compactness (0 – lowest compactness, 3 - highest compactness) without filtering any breath groups.*

Comp.	%V	ΔV	ΔC	varcoV	varcoC	rPVI-V	rPVI-C	nPVI-V	nPVI-C
BrE	40,3	47,2	59,7	59	53,1	50	66,8	61,2	58,9
three	39,8	49,5	70,7	53,6	54,8	53,4	77,4	55,5	59,4
two	39,4	55,2	76,4	57,1	57,0	58,9	82,7	58,1	61,0
one	38,2	54,7	77,2	58,1	57,5	58,1	83,6	58,4	61,9
zero	36,9	54,2	79,5	58,0	59,9	58,8	87,1	59,1	64,4

Tab. 4.11. *Mean rhythm metrics split according to prosodic compactness (0 – lowest degree of compactness, 3 highest degree of compactness) with the short BGs filtered out.*

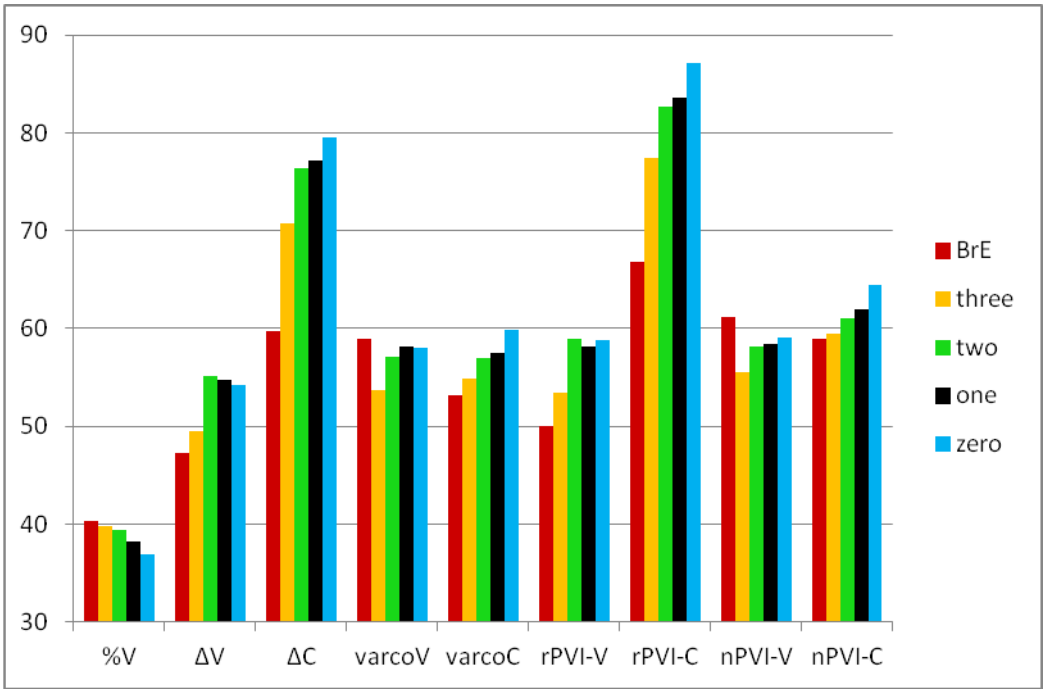


Fig. 4.5. *Rhythm measurements of Vietnamese English based on compactness (0 – lowest degree of compactness; 3 – highest degree of compactness) compared to the British RP standard.*

All BGs were labelled according to their prosodic compactness on the scale from 0 to 3. The decisive factors were the frequency of glottal stops and a subjective sensory assessment through listening. Figure 4.5 shows that compactness, as we have defined it, influences the speaker's rhythm. Except for varcoV and nPVI-V there is either a degree of linearity or at least a significant difference between the value of BrE and the value of the least compact utterances in Vietnamese English. The most compact Vietnamese utterances were closest to the values of British English with the exception of the parameters varcoV and nPVI-V. The difference in parameters ΔC and rPVI-C are again the most prominent in comparison to the other parameters.

4.6 Determining rhythm type

Up to this point we have been dealing with every rhythm parameter separately. Now we are going to address the method to use two parameters at once by means of which

we are able to produce a two-dimensional graph where the speakers are represented by points placed at the intersections of x-values and y-values. The individual speakers should cluster in one area of the graph and their position should give us a better idea on whether a certain dialect or language is more stress-timed or syllable-timed.

We decided to employ two methods based on Volín&Pollák (2009) and Ferragne&Pellegrino (2004). The first uses the figures for %V and ΔC with the presupposition that lower values %V classify the data as more stress-timed and lower values of ΔC as more syllable timed. The latter method employs r-PVI-C and nPVI-V. According to Ferragne&Pellegrino (2004), the higher the values, the more stress-timed the language or dialect.

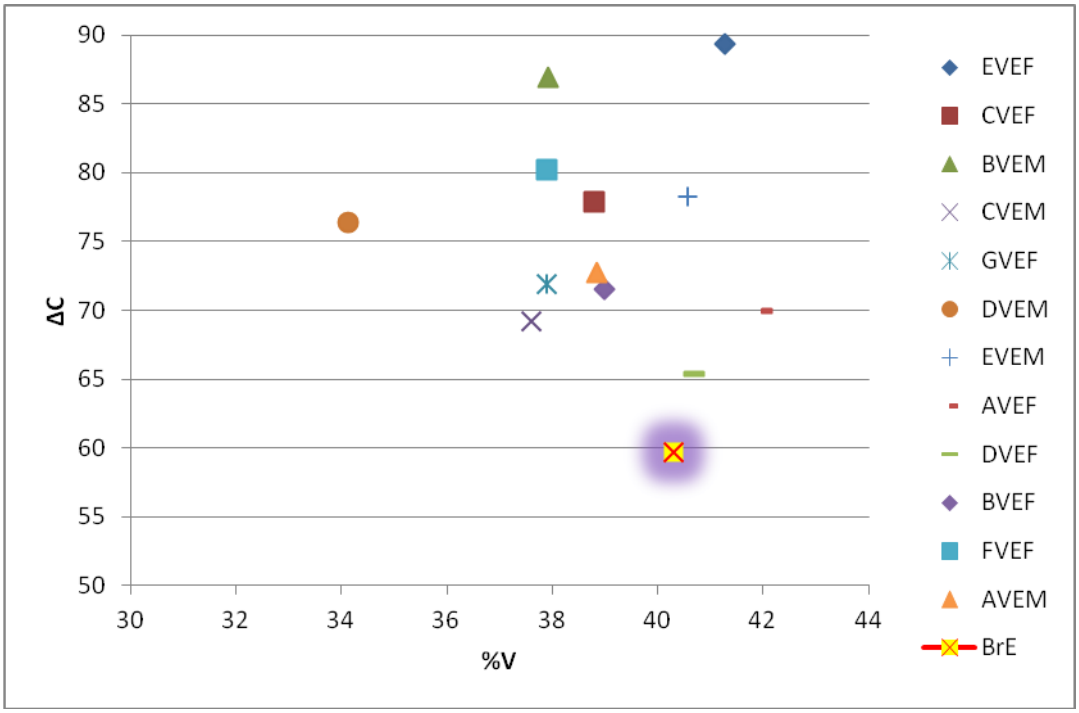


Fig. 4.6. *ΔC and %V values for all the recorded speakers of Vietnamese English. The sign surrounded by shaded area is the average value of the referential BrE RP accent.*

The plot above shows that the values of ΔC are generally higher with Vietnamese English than with RP. Values of %V tend to be similar or slightly lower. Both the

tendencies seem to suggest that Vietnamese English is even more stress-based than the RP standard. Speaker DVEF displays the most similar values to the RP standard among all the recorded speakers. Speakers EVEF, DVEM and BVEM, on the other hand, tend to differ to the greatest extent.

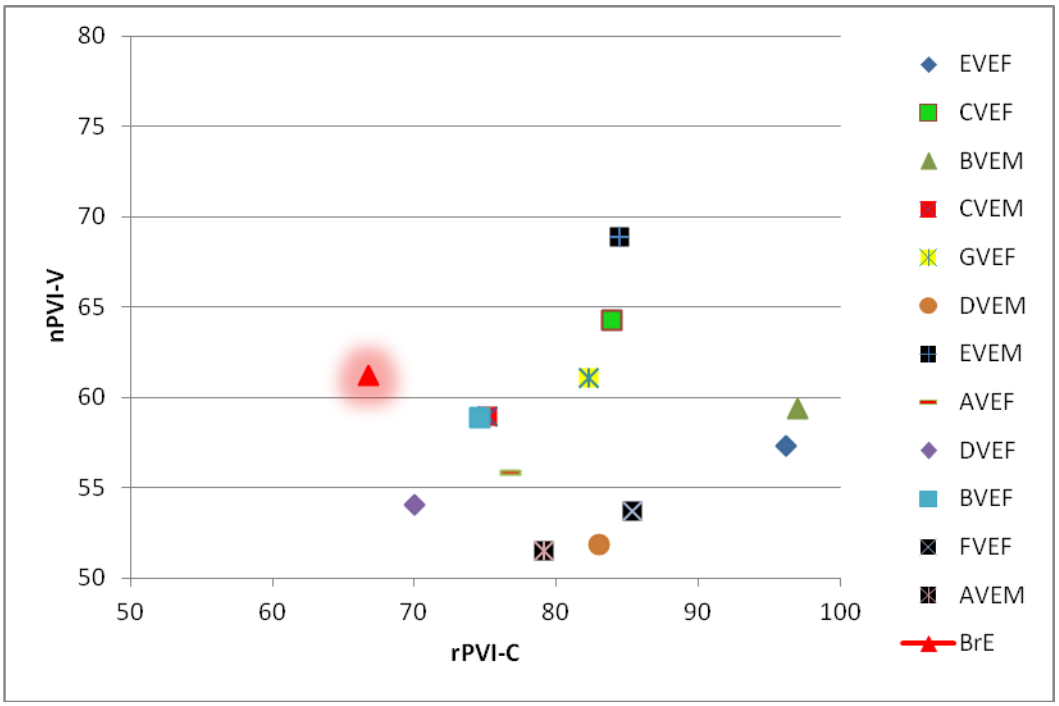


Fig. 4.7. *nPVI-V and rPVI-C values for all the recorded speakers of Vietnamese English. The sign surrounded by shaded area is the average value of the referential BrE RP accent.*

The nPVI-V values for VE tend to be slightly lower than for BrE RP. The difference in rPVI-C is more apparent as we have already discussed in Chapter 4.1. nPVI-C and rPVI-C values also suggest that VE is even more stress-timed than the RP standard. Speakers deviating from the RP standard the most are once again EVEF and BVEM.

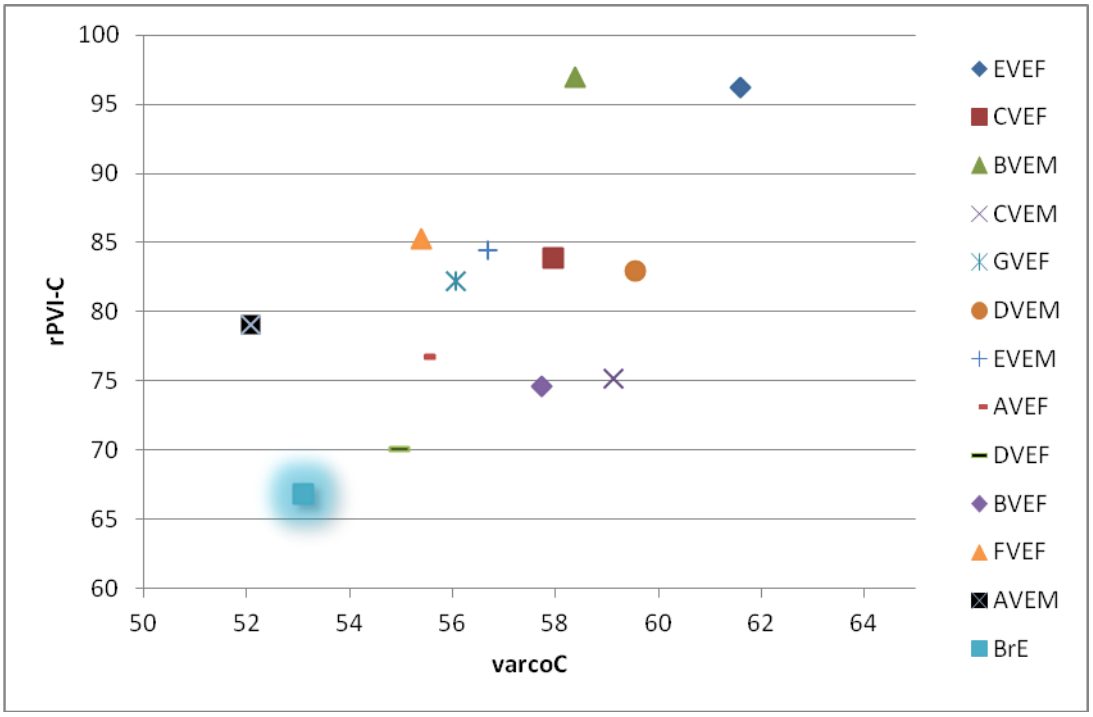


Fig. 4.8. *VarcoC and rPVI-C values for all recorded Vietnamese speakers and the referential British standard.*

The comparison of two consonantal measurements varcoC and rPVI-C confirms the tendency of Vietnamese English towards longer consonantal segments and subsequently towards stress-timing. It must be noted, however, that the parameters of rhythm metrics are merely correlates of rhythm and not rhythmical features as such. Therefore, the results must be interpreted with caution. The graph above further confirms the degree of abnormality for speakers EVEF and BVEM and suggests a necessity to check or even recalculate the values for the two speakers. Speaker DVEF, on the other hand, once again exhibits values closest to the referential British standard.

5. Discussion

5.1 Consonantal measurements

The results show that the discrepancy between vocalic measurements of RP English and Vietnamese is markedly smaller than between consonantal measurements. Vocalic measurements of Vietnamese English are generally slightly lower that of RP. Consonantal measurements of Vietnamese English (especially ΔC and rPVI-C), on the other hand, are higher than in the case of RP. The cause of the consonantal measurements of Vietnamese English being higher can be found in the fact that Vietnamese speakers are generally unable to grasp the technique of linking inherent to the RP English standard. As a result, Vietnamese speakers of English tend to add a glottal stop in front of most word-initial vowels (the less skilled the speaker is the higher tendency to adding glottal stops he or she expresses) rendering the consonantal segments longer than in RP. However, there is a certain possibility that the consonantal prominence in Vietnamese English was artificially and unnaturally amplified by the method of segmentation. Many of the glottal stops displayed a rather long duration (150-300 ms). Whenever a gap of similar duration occurred in front of a consonant, it was labelled as a pause or hesitation (generally gaps longer than 120 ms). Generally, when the speaker's speech rate is slow, the measurements for ΔC are higher. The issue of high ΔC and high frequency of glottal stops is closely connected to the fact that the speakers not only possessed English skills ranging from B1 to low C1 (hence quite far from native like) but also they were not trained in public speaking and presenting news and, moreover, the texts were full of alien place and personal names. It is therefore possible that some of the glottal stops were in fact partially hesitations and they should have been shortened and relabelled. It is nevertheless problematic to state whether or to what extent it was so and whether it would have changed the results significantly. It still seems very likely that the values of consonantal measurements of Vietnamese English would have remained significantly different from the RP standard even after such adjustments in segmentation.

5.2 Rhythm type

High consonantal measurements and the figures 4.6., 4.7. and 4.8. suggest that Vietnamese English is even more stress-timed than the RP standard. Based on the claims about Vietnamese being a syllable-timed language, it seems logical to expect Vietnamese English to be influenced by this characteristic and also display some features of syllable-timing. Nevertheless, as we have mentioned earlier, measurements and calculations of rhythm metrics of the Vietnamese language employing the methodology applied in this thesis have not yet been carried out and therefore it is problematic to make a direct comparison and draw any conclusions.

5.3 Individual deviations

Speakers EVEF and BVEM displayed the highest degree of deviation from the RP standard especially in the consonantal measurements. The difference between the two speakers and the referential values is around 30 units for ΔC as well as rPVI-C, they are even significantly higher than the mode of the rest of Vietnamese speakers (by 10-15 units). The source of this discrepancy is uncertain. There are no clues in linguistic background of the two speakers that would help to identify the cause. When we listen to the recordings of the speakers, it is also problematic to discover a clue explaining the difference in measurements. The speaker BVEM possesses an alien accent that is apparent from the very beginning but the speech of EVEF sounds rather natural from the subjective point of view. The discrepancy could have been caused by mistakes in material processing but if we look through the material, the values of both ΔC and rPVI-C are consistent and there are no BGs with significantly higher values suggesting a mistake in segmentation or calculation.

Speakers DVEF and CVEM, on the other hand, yielded values closest to those of the referential accent and yet they do not seem to sound more native-like than some of the other speakers. In order to be able to comment on this phenomenon, further research would be required, at least a perception test in which a group of respondents would label the recorded speakers from least to most fluent. This finding reflects the recent criticism of rhythm metrics in that they do no measure rhythm as such, let

alone its perceptual attributes, but certain global properties of speech continua which correlate with rhythm types.

6. Conclusion

This thesis attempted to describe the rhythm of Vietnamese English by means of methodology that has been established as default for research in language rhythm across languages. This methodology concerns measuring and comparing rhythm parameters related to vocalic and consonantal duration as described in Chapter 2.2.

Theoretical section of the thesis introduced the phenomena of rhythm in general and the modern means of measuring it. In later sections, it described the basics of the Vietnamese and English language in terms of rhythm as well as in terms of other characteristics such as vocalic and consonantal inventories, tones (for Vietnamese only).

The following chapter on Vietnamese English represents a bridge between the theoretical part and the analysis as it introduces Vietnamese English from the theoretical perspective but, in addition to that, it explains some of the characteristic features of Vietnamese English using examples extracted from the recordings subsequently used for the rhythm analysis.

In terms of vowels, Vietnamese speakers of English tend to experience problems with maintaining desired quantity of English vowels as the Vietnamese language uses vowel quantity as a distinctive feature only marginally. Here it should be noted, however, that English vowels are also distinguished mainly by quality and quality is only a secondary distinctive feature.

In terms of consonants, Vietnamese speakers tend to simplify consonant clusters because there are no consonant clusters present in the Vietnamese language. They often fail to release word-final plosives and fricatives as there are not final fricatives in Vietnamese and the release of the syllable-final plosives is inaudible. Moreover, syllables with final plosives always carry either the rising tone or the rapidly falling tone with strong glottalization. Therefore, many of the words or word-final syllables ending with a fricative or a plosive experience a rise or a glottalized fall in pitch. In Vietnamese, syllables with vowel onset are usually preceded by a glottal stop and so many of the recorded speakers transmitted this feature into their English and they employ glottal stops more often than RP speakers. The last consonant-related feature

comprises deletion of the voiceless fricative [s] at the end of words. This phenomenon could be classified as a subgroup of the insufficient release of word-final fricatives but it needs to be mentioned separately because the Vietnamese speakers sometimes, as if to redeem the failure to pronounce the final [s] properly, add [s] in places where it does not belong, which is a phenomenon that could be classified as a case of hypercorrection.

Chapter 3 describes the means of preparation for the subsequent analysis: how and where was the material gathered, the nature and linguistic background of the recorded speakers and how the material was processed in order to yield the data necessary for the analysis. At the end we postulated five hypotheses and now we present the outcome:

H_A Rhythm Measurements

The measurements of RP and Vietnamese English differ especially with regard to consonants, which is likely to be caused by higher frequency of glottal stops in Vietnamese English.

H_B Gender

Vocalic measurements of female speakers tend to be slightly higher than those of male speakers but according to ANOVA, they bear very little statistical relevance and so no conclusion can be based on the gender difference.

H_C Dysfluencies

Analysis of the data based on the duration of dysfluencies proved to yield no relevant results, which can be explained by the fact that the frequency and duration of dysfluencies was very likely caused by the speakers' unfamiliarity with the texts and there is no correlation between dysfluencies and English skills of the individual speakers.

H_D Compactness

The individual breath groups were assorted according to the degree of prosodic compactness on a scale 0-3 (0 being the least compact, 3 being the most). The value 0-3 was decided by the frequency of glottal stops and a subjective feeling after a careful listening of all the BGs. There proved to be a correlation generally showing that the more compact the BG was the closer the values of the rhythm measurements were to the RP accent.

H_E Rhythm Type

Consonantal measurements being higher subsequently renders Vietnamese English more stress-timed than RP English, which is in conflict with the claim that Vietnamese is a syllable-timed language and hence Vietnamese English is also expected to be more syllable-timed. The issue of whether Vietnamese is a syllable-timed or stress-timed language, however, has not been yet satisfactorily researched and therefore drawing conclusions based on the rhythmical timing of Vietnamese is not yet reliable.

Further Research

In order to ensure the accuracy of the results it would be desirable to check the data yielded by the speakers EVEF and BVEM that deviated the most from the average.

In order to justify the correlation of prosodic compactness, a perception test could be designed, in which a larger number of respondents would label prosodic compactness of the individual speakers.

Most importantly, measurements of rhythm metrics should be carried out on the Vietnamese language. The results of this study would then be better comparable to the nature of the Vietnamese language.

Bibliography

Abercrombie, D. 1967. *Elements of General Phonetics*. Great Britain: Edinburgh University Press.

Boersma, P., Weenink, D. 2011. PRAAT 5.3.03.

Brunelle, M. 2003. *Coarticulation Effects in Northern Vietnamese Tones*. http://aix1.uottawa.ca/~mbrunell/Viet%20coarticulation.pdf

Brunnelle, M. 2009. "Northern and Southern Vietnamese Tone Coarticulation: A Comparative Case Study". *Journal of the Southeast Asian Linguistics Society*.

Brunelle, M. et al. 2012. *Intonation in Northern Vietnamese*. University of Ottawa Press.

Cao Xuân Hạo. 2010. Tiếng Việt, mấy vấn đề ngữ âm, ngữ pháp, ngữ nghĩa (Vietnamese, Selected Problems from Phonetics, Grammar and Semantics). Hà Nội. NXB Giáo dục.

Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge, CUP.

Cunningham, U. 2009. "Phonetic Correlates of Unintelligibility in Vietnamese-accented English". *Proceedings, Fonetik 2009*. University of Stockholm.

Čermák, F. 2004. *Jazyk a jazykověda*. Praha: Karolinum.

Dauer, Rebecca, M. 1987. "Phonetic and Phonological Components of Language Rhythm". *Proceedings 11 ICPhS*. Tallin: IPA. 447-450.

Đoàn Thiện Thuật. 1977. *Ngữ âm tiếng Việt* (Vietnamese Phonetics). Hà Nội: NXB Đại học và trung học chuyên nghiệp.

Ferragne, E., Pellegrino, F. 2004. "Rhythm in Read British English: Interdialect Variability". *Interspeech 10*. 1221-1224.

Ghitza, O., Greenberg, S. 2009. "On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence". *Phonetica 2009; 66*. 113-126.

Grabe, E., Low E. L. 2002. "Durational Variability in Speech and the Rhythm Class Hypothesis". In Gussenhoven, C. and N. Warner. *Papers in Laboratory Phonology 7*. Berlin: Mouton de Gruyter. 515-546.

Hoàng Thị Châu. 1989. *Tiếng Việt trên các miền đất nước – Phương ngữ* (Vietnamese Across the Region – Dialectology). Hà Nội, NXB Khoa học xã hội.

Hombert, J. M., Ohala, J. J. and Ewan, W. G. 1979. "Phonetic Explanations for the Development of Tones". *Language* 55: 37-58.

Hữu Quỳnh - Vương Lộc. 1980. *Khái quát về lịch sử tiếng Việt và ngữ âm tiếng Việt hiện đại* (Brief History of the Vietnamese Language and Modern Vietnamese Phonetics). Hà Nội: NXB Giáo Dục.

Jankovič, M. et al. 2002. Pohledy zblízka: Zvuk, význam, obraz: Poetika literárního díla 20. století. Praha: Torst.

Kirby, J. P. 2011. "Vietnamese (Hanoi Vietnamese)". *Journal of the International Phonetic Association 41(3)*. 381-392.

Liddel, H. G., Scott, R. 1996. *A Greek-English Lexicon*. New York: Oxford University Press.

Machač, P., Skarnitzl, R. 2009. *Fonetická segmentace hlásek*. Praha: Epocha.

Mairano, P., Romano, A. 2011. "Rhythm Metrics for 21 Languages". In *International Congress of Phonetic Sciences*. 1318-1321.

Michaud, A. 2004. "Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese". *Phonetica 61*. 119–146.

Nooteboom, S. 1999. "The Prosody of Speech: Melody and Rhythm". *The Handbook of Phonetic Science*. Blackwell Reference Online.

Palková, Z. 1994. *Fonetika a fonologie češtiny*. Praha: Karolinum.

Pham, A. H. 2003. *Vietnamese Tone, A New Analysis*. New York: Routledge.

Ramus, F. et al. 1999. "Correlates of Linguistic Rhythm in the Speech Signal". *Cognition 73*. 265-292.

Roach, P. 2004. "British English: Received Pronunciation". *Journal of the International Phonetic Association 34 (2)*. 239-245.

Roach, P. 2006. *English Phonetics and Phonology*. 3rd ed. Cambridge: CUP.

Singer, E. 2012. *Vietnamese English*. University of York. Canada.
http://www.yorku.ca/earmstro/asia/Vietnamese_accent.pdf

Slavická, B. 2008. *Praktická fonetika vietnamštiny*. Praha: Karolinum.

Thompson, L. C. 1965. *A Vietnamese Grammar*. Seattle: University of Washington Press.

Trần Trí Dõi. 2011. *Giáo trình lịch sử tiếng Việt* (Textbook to the History of the Vietnamese Language). Hà Nội: NXB giáo dục Việt Nam.

Volín, J. 2005. *IPA-Based Transcription for Czech Students of English*. Prague: Karolinum.

Volín, J., Pollák, P. 2009. "The Dynamic Dimension of the Global Speech-Rhythm Attributes". *In Proceedings of 10th Interspeech*. 1543-1546.

Wagner, P., Dellwo, V. 2004. "Introducing YARD (Yet Another Rhythm Determination) And Re-Introducing Isochrony to Rhythm Research". In *Proceedings of Speech Prosody*. Japan: Nara. 227-230.

White, L. Mattys, S. L. 2007. "Calibrating Rhythm: First Language and Second Language Studies". In *Journal of Phonetics* 35. 501-522.

Yip, M. 2002. *Tone*. Cambridge: CUP.

Czech Résumé

1. Úvod

V úvodní kapitole je nastíněna situace výuky angličtiny v jihovýchodní Asii a problematika nutnosti osvojování přízvuku, který by se co nejvíce přiblížil rodilému mluvčímu. Dále je v této kapitole vymezena oblast zamýšleného výzkumu.

2. Teoretická část

V prvních kapitolách teoretické části je základním způsobem charakterizován rytmus s akcentací na rytmus v komunikaci. Následně je zde popsána moderní metoda výzkumu a měření řečového rytmu za pomoci některých jeho korelátů (%V, ΔV, ΔC, varcoV, varcoC, rPVI-V, rPVI-C, nPVI-V, nPVI-C).

V dalších kapitolách je uvedena stručná charakteristika vietnamštiny, RP (standardní britské angličtiny) a nakonec i vietnamské angličtiny. Charakteristika vietnamské angličtiny se do značné míry zabývá segmentální úrovní a pro podložení uvedených hypotéz jsou využity konkrétní příklady získané při nahrávání materiálu potřebného k analýze rytmu vietnamské angličtiny.

3. Metoda

Tato část popisuje způsob výběru respondentů využitých pro nahrávky a postup nahrávání a zpracování získaného materiálu. Na konci metody jsou uvedeny některé hypotézy, jež budou podrobeny zkoumání v následující sekci.

4. Analýza

V rámci této kapitoly byly vypočítány hodnoty korelátů rytmu (%V, ΔV, ΔC, varcoV, varcoC, rPVI-V, rPVI-C, nPVI-V, nPVI-C) pro vietnamskou angličtinu. Tyto hodnoty byly nejprve porovnány s hodnotami naměřenými pro britský standard.

Druhým krokem bylo vzájemné porovnání hodnot jednotlivých mluvčích vietnamské angličtiny a dále porovnání hodnot s ohledem na pohlaví mluvčích, frekvenci dysfluencí a nakonec s ohledem na prozodickou kompaktnost jednotlivých mluvčích.

Práce dochází k závěru, že vietnamská angličtina vykazuje nepatrně nižší hodnoty vokalických měření oproti britskému standardu. Hodnoty naměřené na souhláskách jsou však u vietnamské angličtiny do značné míry vyšší než u britského standardu. Kategorie věku nevykazuje žádné zásadní rozdíly v naměřených hodnotách. Frekvence dysfluencí taktéž nejspíše nepůsobí měřitelné změny v rytmu. Na druhé straně prozodická kompaktnost rytmus ovlivňuje a kompaktní mluvčí vietnamské angličtiny vykazují hodnoty podobné britskému standardu, kdežto méně kompaktní mluvčí se naopak v hodnotách od britského standardu odchylují více. Rozdíly mezi vietnamskou angličtinou a britským standardem jsou dále zachyceny v bodových grafech porovnávajících vždy dva koreláty rytmu současně.

5. Diskuse

Pátá kapitola se snaží do určité míry objasnit či zhodnotit odlišnosti ve výsledcích a stanovit, co mohlo být příčinou těchto odlišností týkajících se jednotlivých mluvčích, měření souhlásek a určení rytmického typu.

6. Závěr

Závěrečná kapitola shrnuje pozorování uvedená v kapitolách analýza a diskuse. Dále se vrací k hypotézám stanoveným na konci kapitoly metoda a také nastiňuje směr pro další výzkum v oblasti zájmu této práce.